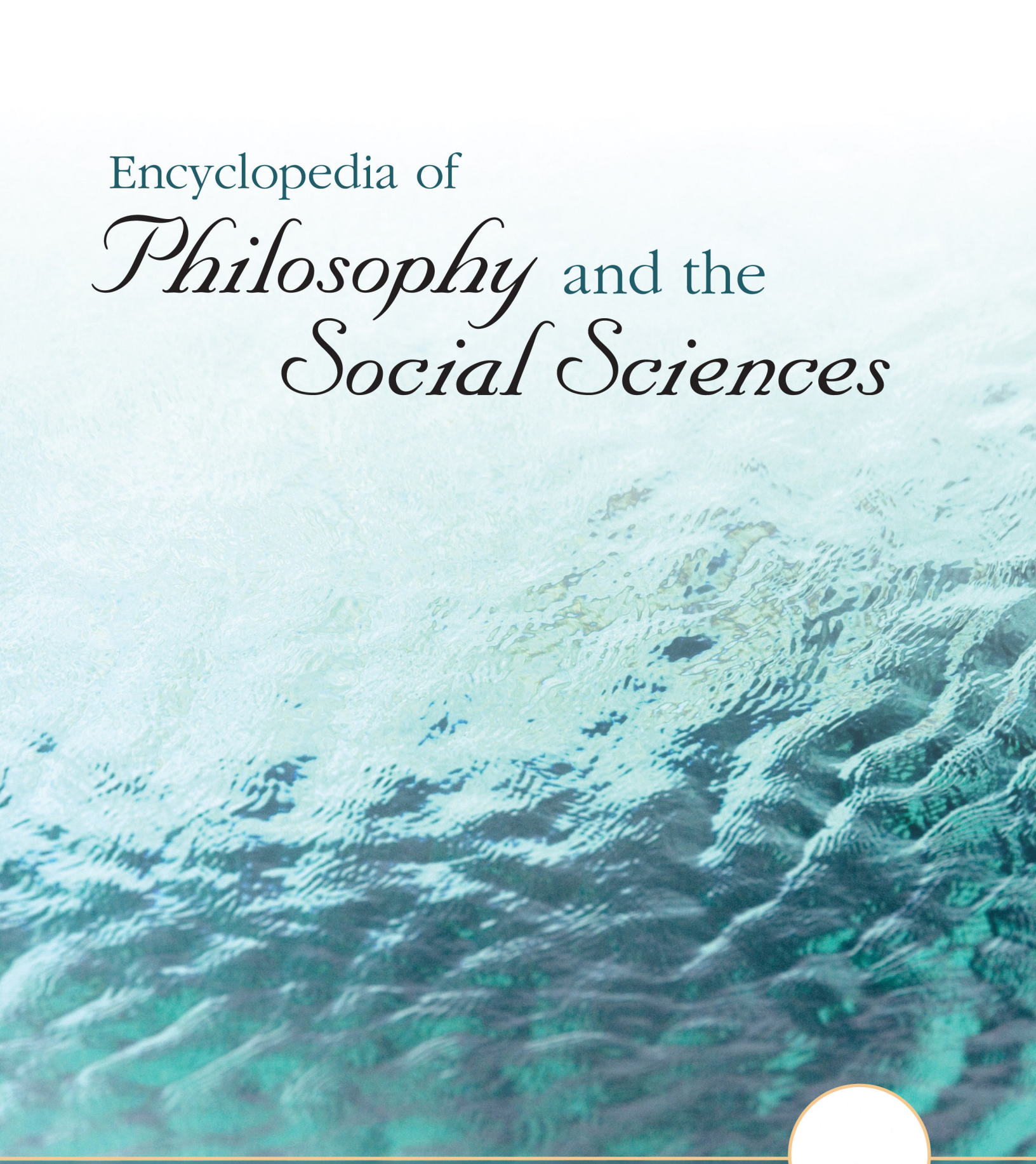
Encyclopedia of *Philosophy* and the *Social Sciences*



Byron Kaldis Editor

Encyclopedia of *Philosophy* and the *Social Sciences*

Editorial Board

Encyclopedia of *Philosophy* and the *Social Sciences*

1

Byron Kaldis

The Hellenic Open University, Greece

Editor

SAGE reference

Los Angeles | London | New Delhi
Singapore | Washington DC

Los Angeles | London | New Delhi
Singapore | Washington DC

FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London, EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Printed in the United States of America.

A catalog record of this book is available from the Library of Congress.

ISBN 978-1-4129-8689-2 (cloth)

Publisher: Rolf A. Janke
Acquisitions Editor: Jim Brace-Thompson
Developmental Editors: Diana E. Axelsen, Sanford Robinson
Production Editor: David C. Felts
Reference Systems Manager: Leticia Gutierrez
Reference Systems Coordinator: Anna Villasenor
Copy Editor: QuADS
Typesetter: Hurix Systems Pvt Ltd.
Proofreaders: Jeff Bryant, Christine Dahlin, Pam Suwinsky
Indexer: Virgil Diodato
Cover Designer: Candice Harman
Marketing Manager: Carmel Schrire

13 14 15 16 17 10 9 8 7 6 5 4 3 2 1

Contents

Volume 1

Volume 2

List of Entries

Reader's Guide

Philosophy, General

A Priori and A Posteriori
Action, Philosophical Theory of
Agency
Alienation: From Philosophy to Sociology
Analytic/Synthetic Distinction
Artificial Intelligence
Behaviorism, Philosophical Conception of
Being-in-the-World
Causation, Philosophical Views of
Cognitive Phenomenology
Complexity
Concepts
Consciousness
Death and Immortality, Philosophical Perspectives
Determinism
Distributed Cognition and Extended-Mind Theory
Embodied Cognition
Emergence
Emotions
Empathy
Empiricism
Encyclopedia
Enlightenment, Critique of
Epistemology
Essentialism
Ethno-Epistemology
Events
Existential Phenomenology and the Social Sciences
Experimental Philosophy
Explanation, Theories of
Feminism: Schools of Thought
Feminist Epistemology
Folk Psychology
Formal Epistemology
Free Will, Philosophical Conceptions of
Given, Myth of the
Hegelianism and Contemporary Epistemology
Hermeneutics, Phenomenology, and Meaning
Holism, in the Philosophy of Language
Idealism
Identity, Personal (Philosophy of)
Inferentialism
Institutions as Moral Persons
Intentionality
Intersubjectivity
Introspection (Philosophical Psychology)
Kinds: Natural Kinds Versus Human Kinds
Knowing-How Versus Knowing-That
Language, Philosophy of
Language-Games and Forms of Life
Laws of Nature
Laws Versus Teleology
Libertarianism, Metaphysical
Life-World
Logical Positivism/Logical Empiricism
Love, Philosophy of
Lying
Mereology: Parts and Wholes
Metaphysics and Science
Mind–Body Relation
Modernity
Modularity of the Mind
Moral Cognitivism
Naturalized Epistemology
Neo-Kantianism
New Wittgensteinians
Nihilism
Nonconceptual Content
Normativity
Objectivity
Philosophes, The
Postmodernism
Pragmatism
Probability

Philosophy of Social Science

Philosophy of Science

Philosophy and Anthropology

Social Ontology and Collective Intentionality

Philosophy and Economics

Marxist Economics
Methodenstreit
Neuroeconomics
Pareto Optimality
Philosophy of Economics, History of
Preference
Reification
Sen's Paretian Liberal
Welfare Economics

Philosophy and History

Annales School
Dialectic, in the History of Philosophy
Dialectic, in the Social Sciences
Herder's Philosophy of History
Historicism
Marxism and Social/Historical Explanation
Narrative in Historical Explanation
Philosophes, The
Philosophy of History
Reification
World-Systems Analysis

Philosophy and Politics

Agnotology, Ignorance, and Uncertainty
Alienation: From Philosophy to Sociology
Analytical Marxism
Behavioralism in Political Science
Capitalism
Common Goods
Communicative Action Theory
Dialectic, in the Social Sciences
Epistemic Approaches to Democracy
Evolutionary Political Science
Governmentality and Regime
Ideology
Institutionalism and Institutional Theory
International Relations, Philosophical and Methodological Debates
Introspection (Philosophical Psychology)
Libertarianism, Political
Marxism and Social/Historical Explanation
Marxist Ethics
Multiculturalism
Neo-Marxism
Neuroscience and Politics
Oppression
Philosophy of Politics, History of
Political Psychology
Public Goods
Public Reason and Justification
Rational Choice and Political Science
Reflective Equilibrium
Reification
Self-Direction and Self-Ownership
Social Contract Theories
World-Systems Analysis

Philosophy and Psychology

Bayesianism, Recent Uses of
Behaviorism in Psychological Explanation
Developmental Psychology
Ego
Embodied Cognition
Existential Psychology
Gestalt Psychology
Intention, Social Psychology of
Personal Identity and Trauma
Phenomenological Schools of Psychology
Philosophical Psychology, History of
Political Psychology
Psychoanalysis, Philosophical Issues in
Schizophrenia: Psychoanalytic, Phenomenological, and Contemporary Philosophical Approaches
Self and the Social Sciences
Therapy, Psychological and Philosophical Issues
Unconscious
Unconscious Social Behavior

Philosophy and Sociology

Actor-Network Theory
Alienation: From Philosophy to Sociology
Analytical Sociology and Social Mechanisms
Communicative Action Theory
Cultural Studies
Death in the Social Sciences
Disciplinarity
Discourse Analysis
Durkheim's Philosophy of Social Science
Economic Sociology
Ethnomethodology
Foucault's Thought
Genealogy
Governmentality and Regime
Group Identity
Habitus
Identity, Social

Philosophy of Action

Cognitive Sciences, Neurosciences, and Social Explanation

Biology and Social Science

Evolution and Social Science

Cooperation, Cultural Evolution of
Cultural Evolution
Evolutionary Ethics
Evolutionary Game Theory and Sociality
Evolutionary Political Science
Evolutionary Psychology
Genetic Indeterminism of Social Action
Human Cultural Niche Construction and the Social Sciences
Neuroethics

Feminism and Social Science

Androcentrism and the Philosophy of Science
Feminism: Schools of Thought
Feminist Critiques of Social Science Applications
Feminist Economics
Feminist Epistemology
Prejudice and Stereotyping
Situational Analysis

Logic and Social Science

Allais Paradox
Bargaining Theory
Bayesianism, Recent Uses of
Coalition Logic
Conventions, Logic of
Decision Theory
Deontic Logic and Agency
Dialogical Logic
Formal Epistemology
Game-Theoretic Modeling
Judgment Aggregation and the Discursive Dilemma
Modal Logic and Intentional Agency
Probability
Social Choice Theory

Sociology of Science

Actor-Network Theory
Situational Analysis
Sociology of Knowledge and Science
Technoscience and Society
Value Neutrality in Science

About the Editor

Byron Kaldis, BA Honours (University of Kent), DPhil (University of Oxford), is Professor of Philosophy and Director of Studies in the School of Humanities at the Hellenic Open University. He has previously held positions at universities in the United Kingdom, United States, and Greece and has recently been a visiting scholar in the departments of philosophy at the University of California, Berkeley and the University of Helsinki. He will be again a visiting scholar at Harvard University and Rutgers University.

He has taught and published in the areas of the philosophy of social sciences, metaphysics and epistemology, history of philosophy and political thought, and the ethics of technoscience. He currently works on the relationship between philosophical issues in cognition and social ontology.

He is the editor of a forthcoming edited volume: *Mind and Society: Cognitive Science Meets the Social Sciences*, to appear in the Synthese Library of Studies in Epistemology, Logic, Methodology, and Philosophy of Science, and co-editor of another forthcoming edited volume titled: *Wealth, Commerce and Philosophy* published by The University of Chicago Press.

Byron Kaldis will be editor of a new academic book series in the philosophy of the social sciences and serves as member of the advisory board on the philosophy of sociality book series at Springer. In 2011, he launched together with colleagues the European Network for the Philosophy of the Social Sciences (ENPOSS).

Contributors

Matthew D. Adler
Duke University

Nicholas Agar
Victoria University of Wellington

Thomas Ågotnes
University of Bergen

Marlies Ahlert
University of Halle-Wittenberg

Stuart C. Aitken
San Diego State University

Martin Albrow
London School of Economics and Political Science

Paul Dragos Aligica
George Mason University

Kristin Andrews
York University

Jonny Anomaly
University of North Carolina at Chapel Hill

David Anzola
University of Surrey

Pat Arneson
Duquesne University

Jakob Arnoldi
Aarhus University

Bill David Ashcroft
University of New South Wales

Richard Badham
Macquarie University

Patrick Baert
University of Cambridge

Aaron B. Ball
University of Chicago

Matthew J. Barker
Concordia University

H. Clark Barrett
University of California, Los Angeles

Tom Beauchamp
Georgetown University

Ulrich Beck
University of Munich and London School of Economics and Political Science

Elisabeth Beck-Gernsheim
NTNU, University of Trondheim (Norway)

Aaron Ben-Ze'ev
University of Haifa

John Bengson
University of Wisconsin–Madison

Gary G. Berntson
The Ohio State University

Michael Biggs
University of Oxford

Ken Binmore
London School of Economics and Political Science

Anders Blok
Copenhagen University

David Bloor
University of Edinburgh

Peter J. Boettke
George Mason University

Samuel Bostaph
University of Dallas

Marcel Boumans
University of Amsterdam

Alban Bouvier
École Normale Supérieure de Paris

Thomas A. Boylan
National University of Ireland, Galway

Geoffrey Brennan
University of North Carolina at Chapel Hill

Susan J. Brison
Dartmouth College

Jan Broersen
Utrecht University

Lara Buchak
University of California, Berkeley

Philip Buckley
McGill University

Mario Bunge
McGill University

Peter J. Burke
University of California, Riverside

David Byrne
Durham University

John T. Cacioppo
University of Chicago

Rafael Capurro
International Center for Information Ethics (ICIE)

Taylor Carman
Barnard College

David W. Carrithers
University of Tennessee at Chattanooga

Nancy Cartwright
University of Durham

Terrell Carver
University of Bristol

Cristiano Castelfranchi
University of Siena

Bikas K. Chakrabarti
Saha Institute of Nuclear Physics

Sara R. Chant
University of Missouri

Anita Chari
University of Oregon

David Christensen
Brown University

Adele E. Clarke
University of California, San Francisco

Randolph Clarke
Florida State University

Robert Clowes
University of Lisbon

Lorraine Code
York University

Finn Collin
University of Copenhagen

George Couvalis
Flinders University

Evard James Cowan
Fairleigh Dickinson University

Mark Cowling
Teesside University

Alice Crary
New School for Social Research

Ann E. Cudd
University of Kansas

Giorgio Hadi Curti
San Diego State University

John Cusbert
Australian National University

Marcello D'Agostino
University of Ferrara

Fred Dallmayr
University of Notre Dame

Arlene Davis
University of North Carolina, School of Medicine

Maximilian de Gaynesford
The University of Reading

Mitchell Dean
University of Newcastle

Florence Delmotte
Facultés universitaires Saint-Louis, Brussels

Robert A. Denemark
University of Delaware

Miraj U. Desai
Fordham University

Willem A. deVries
University of New Hampshire

Paul Dicken
University of Cambridge

David R. Dickens
University of Nevada, Las Vegas

Franz Dietrich
University of Paris-Descartes

Robert Dingwall
Dingwall Enterprises and Nottingham Trent University

Laurent Dobuzinskis
Simon Fraser University

Nigel Dodd
London School of Economics and Political Science

Igor Douven
University of Groningen

John F. Dovidio
Yale University

Keith Dowding
Australian National University

William N. Dunn
University of Pittsburgh

Sven Eliaeson
Uppsala University

Stephen Ellis
University of Oklahoma

Ross B. Emmett
Michigan State University

Pascal Engel
University of Geneva

Edward Erwin
University of Miami

Michael Esfeld
University of Lausanne

David Estlund
Brown Universtiy

Ron Eyerman
Yale University

Don Fallis
University of Arizona

John Feather
Loughborough University

José Luis Ferreira
Universidad Carlos III de Madrid

Roberto Filippi
Birkbeck, University of London

Stephen M. Fiore
University of Central Florida

Michael G. Flaherty
Eckerd College

James R. Flynn
University of Otago

Joseph P. Forgas
University of New South Wales

Peter A. French
Arizona State University

Steven French
University of Leeds

Roger Frie
Simon Fraser University

Carrie Friese
London School of Economics and Political Science

Chris Frith
University College London

Robert Frodeman
University of North Texas

Steve Fuller
University of Warwick

Liane Gabora
University of British Columbia

Wulf Gaertner
University of Osnabrueck

Shaun Gallagher
University of Central Florida

Ana Gantman
New York University

Graeme Garrard
Cardiff University

Ken Gemes
Birkbeck, University of London

Carlos Gershenson
Universidad Nacional Autónoma de México

Bernard Gert
University of North Carolina

Margaret Gilbert
University of California, Irvine

Nigel Gilbert
University of Surrey

Barry K. Gills
Newcastle University

Michael Giudice
York University

Stuart Glennan
Butler University

Natalie Gold
University of Edinburgh

Peter M. Gollwitzer
New York University

Robert M. Gordon
University of Missouri–St. Louis

Mitchell Green
University of Virginia

John D. Greenwood
City College & Graduate Center, CUNY

Till Grüne-Yanoff
Royal Institute of Technology, Stockholm

Armin Grunwald
Karlsruhe Institute of Technology

Francesco Guala
University of Milan

John G. Gunnell
University of California, Davis

Sören Häggqvist
Stockholm University

Alan Hájek
Australian National University

Elżbieta Hałas
University of Warsaw

Ned Hall
Harvard University

Espen Hammer
Temple University

Debora Hammond
Sonoma State University

G. M. Hancock
University of Florida

P. A. Hancock
University of Central Florida

Lynn Hankinson-Nelson
University of Washington

Keith Hart
University of Pretoria, Human Economy Program

Richard F. Hassing
Catholic University of America

Louise C. Hawkley
University of Chicago

Eugene Heath
State University of New York at New Paltz

Marcus Hedahl
Georgetown University

Peter Hedström
University of Oxford

Matti Heinonen
University of Helsinki

Susan Hekman
University of Texas at Arlington

David K. Henderson
University of Nebraska, Lincoln

Stephen Hetherington
University of New South Wales

Vicken Hillis
University of California, Davis

Risto Hilpinen
University of Miami

Frank Hindriks
University of Groningen

Geoffrey M. Hodgson
University of Hertfordshire

Vera Hoffmann-Kolss
Institut für Kognitionswissenschaft Universität Osnabrück

Giora Hon
University of Haifa

Jim Hopkins
Psychoanalysis Unit, University College London

Walter Hopp
Boston University

Mikael Hörnqvist
Columbia University

Paul Hoyningen-Huene
Leibniz University of Hannover

Yan Huang
University of Auckland

Jennifer Hudin
University of California, Berkeley

Wayne Hudson
University of Tasmania and Charles Sturt Universit

Bryce Huebner
Georgetown University

James Hughes
Institute for Ethics and Emerging Technologies

Paul Humphreys
University of Virginia

Ian Hunt
Flinders University

Phil Hutchinson
Manchester Metropolitan University

Marco Iacoboni
University of California, Los Angeles, Brain Mapping Center

Georg G. Iggers
University at Buffalo (SUNY)

Phyllis McKay Illari
University of Kent

Yannis M. Ioannides
Tufts University

Bo Jacobsen
University of Copenhagen

Scott M. James
University of North Carolina, Wilmington

Ian Jarvie
York University

Mark Johnson
University of Oregon

Cory Juhl
University of Texas at Austin

Lutz Kaelber
University of Vermont

Byron Kaldis
The Hellenic Open University

Devorah Kalekin-Fishman
University of Haifa

Annette Karmiloff-Smith
Birkbeck, University of London

Daniel Kelly
Purdue University

Muhammad Ali Khalidi
York University

John F. Kihlstrom
University of California, Berkeley

Harold Kincaid
University of Alabama at Birmingham

Stanley B. Klein
University of California, Santa Barbara

Hartmut Kliemt
Frankfurt School of Finance and Management

Noretta Koertge
Indiana University

Sokratis Koniordos
University of Crete

Roger Koppl
Fairleigh Dickinson University

Arnold Koslow
The Graduate Center, CUNY

Zoltán Kövecses
Eötvös Loránd University, Budapest, Hungary

Michael Krausz
Bryn Mawr College

David B. Kronenfeld
University of California, Riverside

Joel Krueger
University of Copenhagen

Rebecca Kukla
Georgetown University

Martin Kusch
University of Vienna

Hugh Lacey
Swarthmore College

Arto Laitinen
University of Jyväskylä

Kevin N. Laland
University of St Andrews

Mark Lance
Georgetown University

Melissa Lane
Princeton University

Lauren Langman
Loyola University

John Latsis
University of Oxford

Marc Lavoie
University of Ottawa

Alex Law
University of Abertay at Dundee

Clive Lawson
University of Cambridge

Sune Lehmann
Technical University of Denmark

Ernest Lepore
Rutgers University

Neil Levy
Florey Neuroscience Institutes

Paul Lewis
King's College London

Daniel Little
University of Michigan-Dearborn

Omar Lizardo
University of Notre Dame

Sharon A. Lloyd
University of Southern California

Simon Lohse
Leibniz University Hannover

Emiliano Lorini
Institut de Recherche en Informatique de Toulouse

Kirk Ludwig
Indiana University

Sebastian Luft
Marquette University

Steven Lukes
New York University

Steven Luper
Trinity University

Michael E. Lynch
Cornell University

Ronald Macaulay
Pitzer College

Eric MacGilvray
The Ohio State University

Edouard Machery
University of Pittsburgh

Eric Mack
Tulane University

James Maffie
University of Maryland, College Park

Heidi L. Maibom
Carleton University

Rudolf A. Makkreel
Emory University

Lambros Malafouris
University of Oxfrod

Pete Mandik
William Paterson University

Peter T. Manicas
University of Hawaii at Manoa

Leslie Marsh
University of British Columbia

Bill Martin
DePaul University

Nuno Martins
Portuguese Catholic University

Kay Mathiesen
University of Arizona

Tim May
Centre for Sustainable Urban and Regional Futures

Joseph Mazor
Stanford University

Edward McClennen
Syracuse University

Rose McDermott
Brown University

Lee McIntyre
Boston University

Brian P. McLaughlin
Rutgers University

David McLellan
Goldsmiths College, University of London

Alfred Mele
Florida State University

Eduardo Mendieta
Stony Brook University

Mike Mesterton-Gibbons
Florida State University

Janet Metcalfe
Columbia University

Robert G. Meyers
University at Albany, SUNY

Seumas Miller
Charles Sturt University (Canberra)

Ruth G. Millikan
University of Connecticut

Charles W. Mills
Northwestern University

Peter Milne
University of Stirling

Cheryl Misak
University of Toronto

Tariq Modood
University of Bristol

Kristin Renwick Monroe
University of California, Irvine

Michelle Montague
University of Bristol

Leonidas Montes
Universidad Adolfo Ibanez

Dermot Moran
University College Dublin

Peter Murphy
University of Indianapolis

Terry Nardin
National University of Singapore

Thomas Nickles
University of Nevada, Reno

Ilkka Niiniluoto
University of Helsinki

Dmitri Nikulin
New School

Robert Nola
The University of Auckland

Greg J. Norman
University of Chicago

Clare O'Farrell
Queensland University of Technology

Paschal O'Gorman
National University of Ireland, Galway

Gabriele Oettingen
New York University

Gloria Origgi
CNRS

William Outhwaite
Newcastle University

Mark Peacock
York University

Diane Pecher
Erasmus University Rotterdam

Dennis L. Peck
The University of Alabama

Derk Pereboom
Cornell University

B. Guy Peters
University of Pittsburgh

Martin Peterson
Eindhoven University of Techynology

Gualtiero Piccinini
University of Missouri–St. Louis

Massimo Pigliucci
City University of New York, Graduate Center

Sami Pihlström
University of Helsinki

Efraim Podoksik
Hebrew University of Jerusalem

Olga Pombo
University of Lisbon

Bridgette Portman
University of California, Irvine

Garry Potter
Wilfrid Laurier University

Barbara Prainsack
King's College London

Michael Pratt
Queen's University

Dan Priel
York University, Osgoode Hall Law School

Maura Priest
University of California, Irvine

Simon Prosser
University of St Andrews

George Psathas
Boston University

Shahid Rahman
Université de Lille

Julian Reiss
University of Durham

Nicholas Rescher
University of Pittsburgh

George P. Richardson
University at Albany, SUNY

Peter J. Richerson
University of Californai, Davis

Scott Rick
University of Michigan

Mark Risjord
Emory University

Philip Robbins
University of Missouri

Tom Rockmore
Duquesne University

Michael Root
University of Minnesota

David Rosenthal
CUNY Graduate Center

Don Ross
University of Cape Town

Paul A. Roth
University of California, Santa Cruz

Sherrilyn Roush
University of California, Berkeley

Darrell P. Rowbottom
University of Oxford

David F. Ruccio
University of Notre Dame

Robert D. Rupert
University of Colorado, Boulder

Michael Ruse
Florida State University

Federica Russo
University of Kent

Robert Rynasiewicz
The Johns Hopkins University

Mikko Salmela
University of Helsinki

Constantine Sandis
Oxford Brookes University

Louis Sass
Rutgers University

Rapahel Sassower
University of Colorado, Colorado Springs

R. Keith Sawyer
Washington University

Sarah Sawyer
University of Sussex

Margaret Schabas
University of British Columbia

David Lewis Schaefer
College of the Holy Cross

Robert Scharff
University of New Hampshire

Theodore R. Schatzki
University of Kentucky

Marya Schechtman
University of Illinois at Chicago

Samuel Schindler
Aarhus University

Warren Schmaus
Illinois Institute of Technology

Hans Bernhard Schmid
University of Vienna

Frederick F. Schmitt
Indiana University

Stephen F. Schneck
The Catholic University of America

Ralph Schroeder
University of Oxford

Axel Seemann
Bentley University

Evan Selinger
Rochester Institute of Technology

Adam Sennet
University of California, Davis

Sebastian Sequoiah-Grayson
University of Groningen

Russ Shafer-Landau
University of Wisconsin–Madison

Oron Shagrir
Hebrew University of Jerusalem

Wes Sharrock
University of Manchester

Paul Sheehy
Richmond upon Thames College

Giacomo Sillari
Scuola Normale Superiore

Peter Simons
Trinity College Dublin

David Livingstone Smith
University of New England

Declan Smithies
The Ohio State University

Michael Smithson
The Australian National University

Charles T. Snowdon
University of Wisconsin–Madison

Ernest Sosa
Rutgers University

David Spurrett
University of KwaZulu-Natal

Robert Alan Stebbins
University of Calgary

Daniel Steel
Michigan State University

Nico Stehr
Zeppelin University

Diana Strassmann
Rice University

Ron Sun
Rensselaer Polytechnic Institute

Richard Swedberg
Cornell University

James Swindal
Duquesne University

Chris Sykes
Birkbeck College

Justin Sytsma
Eastern Tennessee State University

Arpad Szakolczai
University College Cork

Julia Tanney
University of Kent

Wolfgang Teubert
University of Birmingham

Paul Thagard
University of Waterloo

Mariam Thalos
University of Utah

Georg Theiner
Villanova University

Judith Jarvis Thomson
Massachusetts Institute of Technology

Chris Thornhill
University of Glasgow

Alex Tillas
Heinrich-Heine-Universität

Deborah Perron Tollefsen
University of Memphis

Rosemarie Tong
University of North Carolina at Charlotte

Steve Torrance
University of Sussex

Luca Tummolini
CNR, Institute of Cognitive Sciences and Technologies

Maj Tuomela
University of Helsinki

Charles Turner
Warwick University

Stephen Turner
University of South Florida

John Turri
University of Waterloo

Lars Udehn
Stockholm University

Thomas Uebel
University of Manchester

Peter Vallentyne
University of Missouri

Teun A. van Dijk
Pompeu Fabra University, Barcelona

Donald Phillip Verene
Emory University

Peter Wagner
ICREA and University of Barcelona

David Walton
University of Murcia

Rachel Washburn
Loyola Marymount University

Edward A. Wasserman
University of Iowa

Murray Webster
University of North Carolina at Charlotte

Jeffrey Weeks
London South Bank University

Kate Weiner
University of Manchester

Paul Weirich
University of Missouri

Daniel A. Weiskopf
Georgia State University

Thomas Schwartz Wentzer
Aarhus University

Frederick J. Wertz
Fordham University

David West
Australian National University

Blay Whitby
University of Sussex

Kyle Whyte
Michigan State University

Meredith Williams
The Johns Hopkins University

Jon Williamson
University of Kent

Robert A. Wilson
University of Alberta

Piotr Winkielman
University of California, San Diego

William R. Woodward
University of New Hampshire

K. Brad Wray
State University of New York at Oswego

Steven Yearley
University of Edinburgh

Petri Ylikoski
University of Helsinki

Kei Yoshida
University of Tokyo Center for Philosophy

Julie Zahle
University of Copenhagen

John H. Zammito
Rice University

Jesús P. Zamora Bonilla
Universidad Nacional de Educación a Distancia

Andrzej J. L. Zieleniec
Keele University

Introduction

Philosophy *and/of* the Social Sciences

This Encyclopedia is the first of its kind in bringing together philosophy and the social sciences. As its title suggests, it is not only about the philosophy *of* the social sciences; it is also about the relationship between philosophy *and* the social sciences.

The subject of this encyclopedia is purposefully multi- and interdisciplinary. Knowledge boundaries are both delineated and crossed over in this reference work. The goal is to convey a clear sense of how philosophy looks at the social sciences and to mark out a detailed picture of how exactly the two are interrelated, interwoven at certain times but also differentiated and contrasted at others. The encyclopedia brings forward the ways in which philosophical understanding throws light on the social sciences and, in particular, on their central concepts or key themes; it also explores the ways in which each of the different social-scientific subdisciplines handles such concepts and themes by exhibiting diverse responses to the philosophical analysis of its methods.

Though the theme of this relationship has had its own history and received some academic treatment in the past, there are fresh developments on the current scene. Novel domains are rapidly developing in the area of social ontology and collective intentionality, with discussion of such concepts as shared action, plural subjects, and group mind. New areas of investigation are also emerging at the interface between philosophy and certain modern areas of social-scientific research spawning out of artificial intelligence and cognitive studies and their subfields, which demand a totally new and rather more complex perspective than erstwhile. There are also burgeoning efforts to link logic or its subfields, such as deontic logic, with attempts at regimenting the way human (social or collective) action can be understood.

In addition, it is becoming increasingly apparent that the traditional branches of philosophy (such as philosophy of mind or philosophical psychology) and the social sciences are being linked to each other via recent developments in the area of cognition—for example, research in evolutionary psychology and genetics, primatology, and evolutionary political science, along with neuroscientific studies invading traditional social-scientific fields, has made old-fashioned, rigid divisions between the humanities and social science(s) outmoded. Furthermore, new fields in epistemology, probability theory, and confirmation (e.g., Bayesianism or formal epistemology) represent another area of osmosis between philosophy and the social sciences that has made earlier philosophy of social science rather obsolete.

At the same time, advances in social-scientific research, such as rational choice theory, statistical or stochastic models of decision making, and mathematical modeling of action or game-theoretic approaches, coupled with evolutionary biology or with computer simulation modeling, have in their turn had an impact upon philosophy itself. In this sense, a theme emerging from this work is that there is, at certain loci, a synergistic effect brought about by a process of two-way interaction between philosophy and the social sciences, over and above the one-way study of the social world by means of philosophy.

In this sense, the entries in these volumes cover also fields that are both controversial and on the cutting-edge, thereby foregrounding the central goal

of this project, namely to show the interrelationship between philosophy and the social sciences, especially as it is found in entirely novel niches.

Here is just one case illustrating this tendency of a profoundly changing philosophy of the social sciences: "Homophily" and "social interactions" based on economic models is an example of divergent ways of studying social actors' contact and interaction from within different social disciplines, ranging from sociology to economics, or from novel disciplinary matrices ranging from the analytical study of social mechanisms to complex-networks theory and on to the merging of economic models with physics. It is clear that all such cases raise substantive philosophical issues, thus radically altering the character of the philosophy of the social sciences.

In addition to its novel and multi- and interdisciplinary theme, the encyclopedia's structure is designed so as to best serve as a useful study guide supporting research and instruction at both undergraduate and graduate levels in colleges and universities world wide. It is uniquely placed to serve as both a quick and up-to-date guide for those outside the field(s) and a useful information channel for novel developments and interrelationships for readers who are knowledgeable about one, but not the other, of the two areas here related. This is the added value of this reference work.

To achieve this double benefit for two groups of readers, special attention has been given to two features that are particularly and constantly highlighted. Firstly, the feature of cross-referencing is particularly vital for an encyclopedia with a subject matter that combines two major academic areas, a principal one for the humanities linked with the social sciences. Secondly, because of this distinctive feature and the resultant deeply interdisciplinary nature of the whole project, a Reader's Guide is included to classify entries according to unified themes or subject areas. The inclusion of the cross-references and the Reader's Guide ensures that the encyclopedia is not simply a mirror of achieved knowledge or a catalogue of fossilized dictionary meaning but an active participant contributing to the growth of philosophical knowledge of the social world. Consequently, though the term *reference work* is used in this introduction following established convention, it is important to underline that this encyclopedia is not a mere "reference device" containing dictionary definitions as opposed to encyclopedia entries (to borrow a distinct but related, important idea about a speaker's internal meaning from Gilbert Harman), that is, mere definitions of the meaning of terms as opposed to encyclopedia entries containing facts known to be true descriptions of an outside reality.

An additional feature is the inclusion of entries covering central topics or core historical episodes in both (i.e., central areas of philosophy and of the social sciences), thus assisting readers with no, or limited, knowledge of one of the subject areas to gain a foothold in it by becoming familiar with some of its center points (e.g., entries on Epistemology or on the Philosophy of Language or on Truth, on the side of philosophy, and on Explanation Versus Understanding, on the side of the social sciences). The same is done with classic concepts in the philosophy of science (from which the philosophy of the social sciences has borrowed a lot)—e.g., there are entries on Observation and Theory-Ladenness, on Kuhn, etc. Yet other entries, such as Objectivity or Positivism, etc., straddle all three domains: philosophy, philosophy of science, and philosophy of the social sciences. Finally, a number of entries deal exclusively with the particular "philosophy of" relationships between philosophy and each of the social sciences, as in the philosophy of sociology or the philosophy of history and so forth. Thus, the unique mark of this encyclopedia as serving a neglected educational need is the interrelationship between philosophy and the social sciences and the novel niches thus produced, especially found in fresh ideas and unprecedented hybrid disciplinary areas.

In this way, the encyclopedia serves also a further dual purpose: that of forging a path for the renewal of the philosophy of the social sciences, on the one hand, while on the other helping to establish or promote novel modes of dealing with some of the classic problems where previous attempts have become outmoded or have led to an impasse.

Novel Features

The philosophical searchlight has always been turned on scientific knowledge, but whereas the philosophy of physical and biological sciences is a

well-covered field in terms of textbooks, especially in recent years, the philosophical exploration of the social sciences has remained relatively patchy or partitioned into subfields of social sciences (e.g., philosophy of economics as opposed to philosophy of history, etc.) without a unified and detailed treatment like the one this encyclopedia provides. It is therefore quite important that it includes entries that underline both the novelty of current directions together with the historical tracing of the evolving relationship between its two domains.

Common Concepts

One of the most important aspects of the whole project is the comparative and contrastive exposition of certain core concepts that traditionally receive different or partially compatible or even asymmetrical treatments in philosophy as opposed to social sciences, on the one hand, as well as amongst the various social scientific subdisciplines, on the other. So the reader will encounter an extended number of headwords (from causation to relativism) in double or even triple entries, one dealing with that concept and its role in philosophy juxtaposed to the other covering that concept as it has been understood in the social sciences. In this manner, the reader has a direct view of similarities and differences but also of crucial interrelationships. One is reminded of how one of the classic concepts of social and political philosophy, Jean-Jacques Rousseau's "general will," originated from a purely theological notion (of God's "general will") employed by Antoine Arnauld, Nicolas Malebranche, and others before him.

The encyclopedia therefore follows a novel approach whereby a topic of central and perennial importance is covered jointly by more than one entry and from different angles. This is both useful and pedagogically instructive for students.

Audience

Combining two vast and complex disciplinary areas for the first time, the encyclopedia will be of real benefit to a large and diverse audience. Readers of this encyclopedia will include both undergraduate and graduate students of the humanities and of the social sciences, as well as university professors, researchers, and scholars of both fields. It will also be of benefit to readers unfamiliar with either of the encyclopedia's two domains.

The special aim of this project, as mentioned, is to underline interdisciplinary connections between the humanities and the scientific study of the social world, by drawing analogies, illuminating conceptual kinships, and demarcating interesting contrasts between philosophy and the social sciences, thus for the first time providing readers coming from different disciplinary backgrounds a vista from which to survey novel aspects of the relationships between the humanities and the social sciences.

Thus, the entries in this encyclopedia give readers an opportunity to explore interconnections, clarify commonalities as well as differences or comparative contrasts, discover new fields or ideas of intellectual interest, explore adjacent conceptual zones that may be found to further expand their own disciplinary domains, and also understand better their own academic areas of expertise and the historical provenance of each. The entries are written and arranged with this special aim in view as well as with the goal of introducing students and scholars of a certain scientific discipline to notions, theories, and developments in another with which they are not familiar. The contributors took extra care in writing entries that can successfully accommodate both of these aims.

As was noted above, the overall goal of such a project was to bring together in a fruitful way philosophers and social scientists, even as coauthors of this vast project, and open up, to each, pathways of communication by introducing them to the main elements constituting each other's area. In addition, thanks to the wide coverage of topics and the special selection of entries and their interconnecting lines of cross-reference accomplished in this encyclopedia, undergraduate and graduate university students as well as instructors will have the opportunity to find a wide range of information collected in a single book, rather than seeking it in many different reference works scattered in different places. Students will also be guided by reading the entries on how to enhance their understanding of specific topics, while advanced scholars will be offered the opportunity to trace further interconnections.

Organization of the Encyclopedia and How to Use It

The encyclopedia runs into well over 700,000 words and contains 402 alphabetically arranged entries that range from relatively short ones of c. 1,000 words to mid-sized ones of c. 2,000 words to longer ones of over 3,000 words, and sometimes up to 5,000 words. The latter are meant to be short essays introducing a key concept or major development that requires a substantive analysis extending beyond a brief introduction. Other topics are covered aggregately by more than one entry; in these cases, we have chosen a cluster of shorter entries, each contributing an aspect of the larger topic, while also making sure to provide in addition a longer entry that offers a commanding view of the whole terrain in question. This has been the case mostly with the central and currently burgeoning area of social ontology and collective intentionality, as well as with some of the all-time classics, as it were, of the philosophy of the social sciences, such as the holism/individualism debate or the controversy over the reduction of the social sciences to the natural sciences. In these cases, the topic is discussed directly in entries bearing that title but also in other entries that deal obliquely with the same topic or with that issue (e.g., holism) where and when it had made a central appearance in a particular social science (e.g., psychology or economics). Another example of this approach is the topic of explanation versus understanding or reasons versus causes in action explanation that dovetails with the other two just mentioned, holism/individualism, and reductionism, or with naturalism, dealt in a number of entries.

In all such cases, special care was taken to contrast the issues relevant to the philosophy of the social sciences with the same concerns that had appeared prior to that in the philosophy of the natural sciences or had matured in the latter before they were bequeathed to the social sciences, as is the case with recent attempts of applying Bayesian epistemology, formal-epistemological models, game-theoretic approaches, or computer simulations to the philosophical study of how social explanation does or ought to work. At the same time, we devoted a considerable number of entries to topics in philosophy (e.g., Mind–Body Relation, or Truth) or to the history of philosophy (e.g., Empiricism, Idealism, etc.) and to the philosophy of the social sciences that can be used as background knowledge for readers and also as pointers for further interconnections.

Cross-References

This encyclopedia emphasizes the interconnectedness of learning and of knowledge in general; therefore, special care was taken to identify cross-references and further readings with an eye to such interconnectedness. This resembles a quasi-coherentist principle of encyclopedic knowledge, leading the reader from one headword to another (not to be confused with the distinction Harman attacked, as mentioned previously), or what can be called the "Hobbesian principle": coming across the demonstration of the Pythagorean theorem by chance, when glancing over a page in a copy of Euclid's Elements open before him, Thomas Hobbes was so astonished by the theorem that he determined to convince himself of its truth by tracing the interconnecting steps leading from one proposition or partial proof to another.

We strongly recommend that readers seek out the suggested cross-references listed in the "See also" section at the end of each entry. These cross-references have for the first time a double goal: to point to associated topics, further links, and related issues of kindred nature, or also to a concept or topic from the "other side" as it were—that is, either from philosophy or from the social sciences (e.g., Death, Space, Complexity, Time), on the one hand, or even to a totally contrasting notion, on the other hand (i.e., on certain occasions we include under "See also" an entry that contains a contrasting topic so that the reader is led to consult that as well). The latter device is of importance for it serves the overall purpose of this work, namely to pinpoint not only interconnections but also differences between philosophy and the social sciences.

The Reader's Guide

To guide the reader to clusters of interwoven items and kindred areas, all the entries have been listed thematically in the Reader's Guide, which

follows the List of Entries in the front matter. Each entry has been listed in at least one of the following seventeen subject categories:

- Philosophy, General
- Philosophy of Social Science
- Philosophy of Science
- Social Ontology and Collective Intentionality
- Philosophy and Anthropology
- Philosophy and Economics
- Philosophy and History
- Philosophy and Politics
- Philosophy and Psychology
- Philosophy and Sociology
- Philosophy of Action
- Cognitive Sciences, Neurosciences, and Social Explanation
- Biology and Social Science
- Evolution and Social Science
- Feminism and Social Science
- Logic and Social Science
- Sociology of Science

This listing is itself a significant contribution to the dynamically developing philosophy of the social sciences, since the taxonomies presented here (e.g., Biology and Social Science or Evolution and Social Science or Logic and Social Science) can themselves be contestable disciplinary divisions that play a significant role. As we said, quite a few of the entries are classified under more than one of the above themes, given that the essential feature of this reference work is the interconnections between concepts and issues appearing both in philosophy and in the social sciences.

Further Readings

Finally, each entry ends with a helpful list of Further Readings that are designed to include representative work by thinkers mentioned, or items referred to, in the body of the entry. These listings will enable readers to locate the principal texts in question and to find additional sources through which they can pursue further their study of the specific topic.

Acknowledgments

I would like to thank all the authors for their high-quality contributions; the members of the Editorial Board—Ann Cudd, Margaret Gilbert, Ian Jarvie, Tony Lawson, Philip Pettit, John Searle, Raimo Tuomela, and Stephen Turner—for embracing the idea right from its inception, as well as for their support; and all the people at SAGE who made it possible for this project to materialize: in particular, the development editors, Diana Axelsen and Sanford Robinson, for the substantive and knowledgeable help they provided throughout but also for their encouragement and patience; the SAGE reference tracking system coordinators, Anna Villasenor and Laura Notton, for their flawless work and graceful assistance at all times; the production editor, David Felts, and the copyeditors and proofreaders who went over the whole manuscript with meticulous care and professional expertise; the senior marketing manager, Carmel Schrire, for her efforts in promoting the final product; and, of course, Jim Brace-Thompson, the senior acquisitions editor, who believed in this project from the outset. Special thanks go to my friend and colleague Leslie Marsh for his valuable input at the start.

Byron Kaldis

A Priori and A Posteriori

According to a distinction codified by Immanuel Kant and broadly agreed upon in contemporary epistemology, epistemically justified beliefs are divided into two kinds. Those that are *a posteriori* justified are justified by experience, most importantly by empirical observation using the five external senses. For example, your belief that there is now English text in front of you is a posteriori justified since it is based on your current visual experience. *A priori* justified beliefs are nonexperientially justified; they are justified by powers such as intuition and rational insight. Your belief that $7 + 5 = 12$, for example, is a priori justified if it is based on your mental grasp of this sum or a mental calculation. Understanding this important distinction helps one identify the respective roles that empirical evidence and armchair thinking play in justifying beliefs, including beliefs in the social sciences. This entry further clarifies this distinction and then reviews the main kinds of a posteriori and a priori justified beliefs. Along the way, some examples of a priori and a posteriori justified beliefs in the social sciences will be highlighted.

A posteriori justification and a priori justification are species of epistemic justification. A belief is epistemically justified if it is well aimed at the truth. Beliefs may target practical or moral goals; epistemic justification, though, is solely concerned with whether a belief is well aimed at the truth. Being well aimed at the truth is significantly different from being true. For example, a wild guess that happens to be correct is not well aimed at the truth, so it is true but not justified. Conversely, an 18th-century scientist's belief in Newtonian mechanics was justified despite being false. Since the general nature of epistemic justification allows for justified false beliefs, some such beliefs are a priori justified. Someone, for example, who performs a mental calculation that leads her to believe that $67 \times 67 = 4{,}549$ might have an a priori justified belief (since it was arrived at using a method available from the armchair), even though that belief is false.

It is important to see how the a priori/a posteriori distinction differs from the innate/acquired distinction. Kant highlighted this key difference. While a person might need to undergo some specific experiences in order to furnish his mind with the concepts he needs to form some belief, the justification for that belief might nonetheless be a priori. For example, suppose (as is plausible) that humans are not innately endowed with concepts of numbers or concepts of arithmetic operations; instead, they acquire these concepts from their environment. In this case, a person's belief that $67 \times 67 = 4549$ depends on experience to enable him to have thoughts about multiplication, equality, 67, and 4549. However, if his evidence for this belief consists in a mental calculation that he performs, that belief is a priori justified.

Let's turn next to classification. Justified beliefs based on observation using our five external senses are obviously a posteriori justified. Though these beliefs are clearly central to the social sciences, they probably do not exhaust the justified beliefs of social scientists. Other justified beliefs originate from sources that have been traditionally classified as a posteriori: testimony, introspection, memory, and

the internal senses that report our bodily states. Of these, testimony probably plays the most important role in the social sciences—for example, in social anthropology, criminology, and economics.

Most figures in the history of epistemology and contemporary epistemology endorse some a priori justified beliefs. They often argue from cases, pointing to justified beliefs in the areas of mathematics, logic, and ethics; justified beliefs in analytic and definitional truths (e.g., all bachelors are male); and justified beliefs that arise from reflecting on thought experiments. Justified beliefs of each of these types show up in the social sciences. Some social sciences, like economics, make heavy use of mathematics. Deductive and inductive rules of inference from logic are used throughout the social sciences whenever a conclusion is inferred from evidence. Whether ethical claims have a legitimate place in social-scientific inquiry is a vexed issue; but if they do, this is another place for a priori justified beliefs. Beliefs in analytic and definitional truths are at work when social scientists employ their specialized terminologies. And if there are social-scientific thought experiments that yield justified beliefs (e.g., in economics), then this too might be a place for a priori justified beliefs.

Radical empiricists like John Stuart Mill argue that there are no a priori justified beliefs. They claim instead that only a posteriori sources can justify beliefs. They have two options for responding to any alleged a priori justified belief: Either argue that it is not justified at all or argue that it is a posteriori justified. Those who oppose these empiricists and claim that there are a priori justified beliefs often offer two kinds of arguments. One invokes the nonempirical practices of mathematicians, logicians, and ethicists, pointing out, for example, that university mathematics departments do not have labs. The other appeal is to beliefs that some claim are necessarily (as opposed to contingently) true. Proponents of a priori justified beliefs, like Kant, maintain that the justification for these beliefs cannot be construed as a posteriori, since empirical observation only tells us how things happen to be, not how they must be. These beliefs are therefore justified a priori.

Peter Murphy

See also Analytic/Synthetic Distinction; Empiricism; Experimental Philosophy; Logical Positivism/Logical Empiricism; Naturalized Epistemology; Thought Experiments

Further Readings

Casullo, A. (2003). *A priori justification*. New York, NY: Oxford University Press.

Kant, I. (1965). *Critique of pure reason* (N. K. Smith, Trans.). New York, NY: St. Martin. (Original work published 1781)

Kripke, S. A. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.

Moser, P. K. (Ed.). (1987). *A priori knowledge*. New York, NY: Oxford University Press.

Abduction and Inference to the Best Explanation

Abduction, or—as it is more commonly called nowadays—inference to the best explanation, is one of three major types of inference, the other two being deduction and induction. The distinction between deduction, on the one hand, and induction and abduction, on the other, corresponds to the distinction between necessary and nonnecessary inferences. In deductive inferences, what is inferred is necessarily true if the premises from which it is inferred are true. This is not so for the other two types of inference. Broadly speaking, in inductive inferences a conclusion is reached on the basis of statistical information. This type of inference is at work when a doctor concludes that a patient has a certain ailment given the patient's symptoms and given that almost all people who have those symptoms have the ailment. What singles abduction out among the nonnecessary inferences is that in abductive inferences a conclusion is reached on the basis of explanatory considerations, as when we conclude that a friend has missed the train since that best explains why she is late.

The term *abduction* was coined by the American pragmatist philosopher Charles Sanders Peirce in the context of his work on the logic of science. For him, the term did not quite mean what we mean by it nowadays. The main difference between his conception and the modern one is that for Peirce, abduction had its proper place in the so-called context of discovery, the stage of inquiry in which we try to generate hypotheses, which may later be assessed. In particular, he saw abduction as a guided process of forming hypotheses, where explanatory considerations serve as the main guide. Deduction

and induction, then, come into play in the so-called context of justification, in which we are concerned with the assessment of those hypotheses: Deduction helps derive testable consequences from them, and induction finally helps us reach a verdict on them.

By contrast, in the modern conception of abduction, this type of inference belongs in the context of justification as well. The idea is that we may be warranted to accept a hypothesis on the basis of its explanatory power, as the standard textbook formulation of abduction brings out:

> ABD Given evidence E and candidate explanations $H_1, \ldots, H_n$ of E, accept the H_i that best explains E.

For example, given the evidence that our friend is late, and given a number of possible explanations for why she is late, we conclude that she missed the train, because that best explains her being late.

Not only is abduction common in everyday life, philosophers of science have argued that it is a cornerstone of scientific methodology. Often, scientists are confronted by the fact that a number of rival hypotheses that they are considering are all consistent with the available data. The idea is that where the data alone are insufficient to warrant a choice between such hypotheses, an appeal to explanatory power may break the deadlock. Of two or more hypotheses that are consistent with the data, one may still stand out as best explaining those data, given that consistency with the data is necessary but not sufficient for explaining those data.

While it is generally agreed that, as a psychological matter of fact, abduction finds frequent employment in both our everyday and our scientific reasoning, the status of abduction is controversial. A major worry is that the best explanation will always only be the best of the candidate explanations that we have been able to conceive. Given that we generally will not be able to think of all possible explanations for the evidence at hand, and do not have the guarantee that we will at least be able to think of the correct explanation, the explanation we pick as being the best may well be the best of a bad lot (as was famously argued by Bas van Fraassen).

It has been argued in response that this worry arises only due to the rather simplistic way in which ABD and kindred textbook presentations explicate the idea of abduction. Specifically, the worry arises due to the fact that ABD gives license to an absolute conclusion—that a given hypothesis is true—on the basis of a comparative premise, namely, that that particular hypothesis is the best explanation of the evidence relative to the other hypotheses available. To undo this asymmetry, we can either have the rule require an absolute premise—for instance, that the hypothesis whose truth is inferred not only be the best of the available potential explanations but also be good enough—or have it sanction, given a comparative premise, only a comparative conclusion: For instance, the best explanation is closer to the truth than any of the other available candidate explanations.

Even if sufficiently sophisticated versions of the rule of abduction escape objections like the one mentioned, there is still the question of whether application of the rule is rational. More exactly, the question has been raised whether following the rule has any tendency to increase the number of truths among our beliefs. There are several ways to argue that the rule does have this tendency. The best-known argument for the claim that abduction is truth conducive is an empirical argument developed by Richard Boyd.

The argument starts by underlining the theory dependency of scientific methodology, which comprises methods for designing experiments, assessing data, choosing between rival hypotheses, and so on. For instance, in considering possible confounding factors from which an experimental setup has to be shielded, scientists draw heavily on already accepted theories. The argument next calls attention to the apparent reliability of this methodology, which, after all, has yielded, and continues to yield, impressively accurate theories. In particular, by relying on this methodology, scientists have for some time now been able to find ever more instrumentally adequate theories. Boyd then argues that the reliability of scientific methodology is best explained by assuming that the theories on which it relies are at least approximately true. From this and from the fact that most of these theories were arrived at chiefly by abductive reasoning, he concludes that abduction must be a reliable rule of inference.

Igor Douven

See also Deduction; Explanation, Theories of; Induction and Confirmation; Instrumentalism of Scientific Theories and Constructive Empiricism; Observation and Theory-Ladenness; Probability; Scientific Method

Further Readings

Boyd, R. (1981). Scientific realism and naturalistic epistemology. In P. Asquith and R. Giere (Eds.), *PSA 1980* (Vol. 2, pp. 613–662). East Lansing, MI: Philosophy of Science Association.

Douven, I. (1999). Inference to the best explanation made coherent. *Philosophy of Science, 66,* 424–435.

Frankfurt, H. (1958). Peirce's notion of abduction. *Journal of Philosophy, 55,* 593–596.

Harman, G. (1965). The inference to the best explanation. *Philosophical Review, 74,* 88–95.

Kuipers, T. (2000). *From instrumentalism to constructive realism.* Dordrecht, Netherlands: Kluwer Academic.

Lipton, P. (2004). *Inference to the best explanation* (2nd ed.). London, England: Routledge.

van Fraassen, B. (1989). *Laws and symmetry.* Oxford, England: Clarendon Press.

Action, Philosophical Theory of

The philosophical theory of action relates to issues as diverse as those of agency, autonomy, criminal liability, Decision Theory, human nature, practical reasoning, free will, mental acts, individual and collective responsibility, intention, motivation and explanation, rationality, speech acts, addiction, compulsion, and weakness of will. So while it is a distinct area of enquiry in its own right, many philosophical views on action have been expressed from within various fields, including philosophy of mind, ethics, metaphysics, legal and political philosophy, philosophy of history, philosophy of language, philosophy of science, behavioral science, and cognitive ethology. It is thus central to the philosophy of the social sciences.

Ontology, Individuation, and Agency

Action theory has also been central to modern philosophy of mind, ever since René Descartes's contemporaries first criticized the account of *mental causation* suggested by his substance dualism by asking when, where, and how the nonphysical mind caused bodily movement (i.e., the immaterial substance *mind* caused the material substance *body* to act, and vice versa, a thesis called psychophysical interaction). Descartes famously located the crucial psychophysical event in the pineal gland, but he may have done better to maintain that when A causes B it is ontologically promiscuous to postulate some third event of A's causing B, itself in need of a further cause, ad infinitum.

While some philosophers use the terms *action* and *behavior* interchangeably, others reserve the former for behavior that is intentional and/or voluntary (at least under some description). Actions are frequently identified by theorists as events and/or processes; however, there is much dispute over *which* events actions are to be identified with. Donald Davidson, for example, maintains that actions are identical to certain movements of our bodies. Yet, as Jennifer Hornsby has cautioned, we must not conflate the *act of moving* one's body with (mere) bodily movement (the term may be used in both a transitive and an intransitive sense). Hornsby further distinguishes between the thing one did and (the event of) one's doing it.

How one individuates action depends on one's understanding of what it is for an action to be *basic,* in particular whether the difference between basic and nonbasic action is a matter of *kind, degree,* or *description.* Arthur C. Danto introduced the term *basic action* with the aim of picking out the subset actions that have not been caused to happen (at least not by the agent doing anything else first). He gives the example of his moving his arm without doing anything to cause it to move. Yet on this account, *almost all* actions would count as basic. As Annette Baier has noted, the search for basic action is a hunt for the most manageable and/or minimal cases of action. Without an account of what it is for an action to be basic, we can have no firm conception of what agency consists in.

Alvin Goldman has criticized Danto for confusing *causation* with *causal generation,* to be further distinguished from conventional generation, simple generation, and augmentation generation. The problem of specifying what we mean by "basic" remains even after we have settled for a specific causal, teleological, or meditative relation. Baier accordingly divides actions into at least eight kinds of "basicness": causally basic, instrumentally basic, conventionally basic, ontologically basic, logically basic, genetically basic, ease basic, and isolation basic. To this we might add epistemically basic, and no doubt further categories may be imagined.

Whatever notion of basicness we stipulate, it is tempting to think of qualifying actions as token members of a *class.* So understood, basicness is an *absolute* property (such as that of being saturated) and not a relative one (as is tallness). Such a conception nonetheless allows for varying *degrees* of

approximation: My act of operating the pump may appear less basic than that of moving my arm but more so than that of replenishing the water supply (itself less basic than that of poisoning the inhabitants). But how distinct are these four acts?

Irving Thalberg has labeled as *reductive unifiers* those who side with Elizabeth Anscombe in claiming that we here have one action with several different descriptions. This is because they maintain that being basic is a matter of *description,* not kind. The most influential of these claimants, Davidson, maintains that *all* actions are events that are basic under-some-description (one among many descriptions under which they will also be intentional). By contrast, *pluralists* or *multipliers,* such as Goldman and Judith Jarvis Thomson, claim that each of the above descriptions picks out a different action, only one of which can be the most basic. Thalberg's own position, *nonreductive unification,* rejects the choice between identity and independence in favor of a part–whole relation, according to which the event of my replenishing the water supply, for instance, *includes*—but is not identical to—the event of my operating the pump. But whether the two events are numerically distinct is a moot point, and we may prefer to follow Hornsby's talk of *identifiers* and *differentiators.* Moreover, we should be wary of equating the individuation of events with that of things we *do,* for there may be one event of my doing several distinct things. Still, we can at least sometimes offer multiple descriptions of one and the same action: Oedipus does not kill Laius *and* his father, though it is true that he *strikes* Laius and that he kills his father.

Suppose that Bob Marley shot the sheriff at time *t*1 but the sheriff only died at a later time, *t*3, before which—at time *t*2—Marley records his famous song. Did Marley kill the sheriff before or after recording his song? (He certainly didn't do it during the recording.) It would be as implausible to claim (with Goldman and Thomson) that Marley did not kill the sheriff until *t*3—after he had left the scene of the crime and was back in the recording studio—as it would be to follow Davidson and Hornsby in maintaining that he killed the sheriff at *t*1, before the sheriff died.

Jonathan Bennett has objected that the implausibility of the latter claim is not metaphysical but linguistic. We do not *call* a woman a mother before she has any children, yet we may, after the birth or adoption of her first child, legitimately speak of what the child's *mother* did before she had any children. Similarly (or so the argument goes), while we cannot at *t*1 (when the sheriff was still alive) truthfully say that Marley has killed the sheriff, at *t*3 (when the sheriff is dead) it becomes perfectly acceptable to talk of the *killing* having occurred at *t*1. Yet if it proves anything, the analogy seems to prove the opposite, for we do not imply that the woman was a mother *before* the advent of the child, and we likewise cannot plausibly state at *t*3 that Marley had killed the sheriff at *t*1, even though it is correct to say that the person who shot him then was indeed (to be) the sheriff's *killer.* A variant of this view claims that events acquire properties over time, much as physical objects, like people, acquire properties over time: Just as a deceased man may, at the time of my birth, acquire the property of being my great grandfather, so Marley's act of shooting acquires the property of being a killing once the sheriff dies. But while we know what it is for the event of Alekos Sandis's birth to acquire later the property of being the event of his great grandfather's birth (through what Peter Geach has called a "Cambridge change," that is, by virtue merely of my subsequent birth, without any change effected on him), it is less clear how a shooting can become a killing without *itself* changing. This is because killings are *causings* of death, whereas shootings are at most a *cause.*

We may call both causings and causes "events"; it would be absurd to think of the causing of an event as something that could itself be brought about (though it may be that *in* raising my arm I make it the case that the event of my raising my arm occurs). Causings may always be spatio-temporally in varying degrees of fine-graininess. To always insist on the same precision of location with regard to them leads to the absurdity of insisting that all actions must have a smaller spatial location than, say, that of an average-sized dining table. If Marley shot the sheriff in March 1973 (before recording his song about it in April 1973) and the sheriff (unlike the deputy) does not die until November 1973 (after the hit record was released), then Marley killed the sheriff in (no particular month of) 1973. The impossibility of locating any given event more precisely than in a certain minute, hour, day, week, month, season, year, decade, or century does not imply continuous duration throughout that period (think of cricket matches).

Volition and the Will

Most rationalists and empiricists believed in the existence of conative *mental acts* of "the will" ("conative"—from *conatio,* a tendency or desire to

attempt to do something—is standardly contrasted to "cognitive" and is used to refer to the volitional part or desiring aspect—i.e., "wanting to"—of a mental act or process). Volitionism comes in numerous forms, stemming from different understandings of the relation between volition and "bodily movement."

One popular form may be labeled "volition *as* action." According to this Cartesian view, all actions are mental "acts of will" that cause our bodies to move. Thus, for example, H. A. Prichard maintains that acting is a successful form of willing: When I move my hand, the movement of my hand, although an effect of my action, is not itself an action or even part of one. A recent variant of this position defended in Hornsby's early work is the view that all actions are *tryings*. One way of reaching this conclusion is by means of the argument from failure: Since it is always possible to fail to do what we set out to, the most *basic* thing we ever do is attempt or *try* to act. Everything else we do is a *consequence* of such basic acts. Such views can have important consequences in ethics. Prichard concluded that while we are morally obligated to *try (*our best) to execute our obligations, we cannot be obligated to succeed.

A second form of volitionism claims not that actions are volitions but that an action is a (mere) bodily movement *caused* by a volition. On this view (espoused by empiricists such as John Locke and David Hume), actions are *voluntary movements,* defined further as movements caused by a volition: We will to do something, and the act (viz., the bodily movement) follows (note the aforementioned ambiguity of the term *bodily movement*).

A third form of volitionism claims that action is *sometimes* a volition and *sometimes* a bodily movement caused by a volition. Proponents of this view (including Thomas Hobbes, Thomas Reid, and Jeremy Bentham) hold that we must therefore distinguish between two *kinds* of action, *internal* and *external,* the former being "acts of mind," such as that of intending, and the latter corporal "acts of the body," such as that of moving one's fist. A fourth identifies action not with one thing but a series of two: a mental event followed by its physical effect. A related account identifies action with the process or event of a volition's *causing* a bodily movement.

Volitionism suffered a serious blow at the hands of Ludwig Wittgenstein, who, questioning what sort of thing a "willing" was meant to be (is it something we can do intentionally? voluntarily?), argued that whether or not an action counts as either voluntary or intentional does not depend upon the causality of some "inner" event of volition or intention but on the wider context in which the action was performed (e.g., was it coerced or performed under duress?). These suggestions were subsequently elaborated upon and developed in various directions by his students. The suggestion that action is the *expression* (as opposed to effect) of one's beliefs and desires brought Wittgensteinians in an unexpected allegiance with Hegelians and Collingwoodians (i.e., British idealists like R. G. Collingwood), which continues to this day (e.g., in the Pittsburgh school).

In opposition to this behavioristic zeitgeist, Peter Geach has argued that while we often distinguish between *thinking* and *acting,* many cases of thinking, calculating, and judging are episodic cognitive mental acts, which we often perform intentionally, and for reasons, and which have duration and phenomenology. In a similar vein, J. L. Austin reminds us that utterances can also be understood as *speech acts* of describing. Some *illocutionary* speech acts are *performative utterances*—that is, performances of the very actions they describe, for we do numerous things "with words," for example, make promises, get married, name ships, deny accusations, and give orders. We also perform *perlocutionary* speech acts of convincing, scaring, emphasizing, persuading, annoying, and so on.

Reasons, Motivation, and Explanation

Intentional action frequently overlaps with action performed for reasons, though both a priori and experimental philosophy have recently challenged the assumption that (a) intentional actions are always performed for reasons and (b) actions performed for reasons are always intentional.

In arguing that the primary reason for an intentional action is its cause, Davidson attacked the received Wittgensteinian view, replacing it with an influential version of the Humean theory of motivation (though whether David Hume himself held this view is another matter). Humeanism became the orthodoxy in analytic philosophy for the second half of the 20th century. In recent years, however, it has recently faced a new wave of resistance from defenders of agent causation of one form or another, as well as from those who maintain that it is a mistake to view the reasons for which we act as psychological states.

These and other debates are frequently couched in terms of "motivating reasons." This has misled philosophers to write as if questions about what *motivates* us to act are identical to questions about the *reasons* for which we act. But although we can be motivated by greed, greed need not be a consideration we act upon, and as detective fiction reveals, one may have a motive to do something without being in the least inclined to act upon it.

Additional questions, such as whether facts can be said to cause anything and how we should understand the distinction between causal relations and causal explanations, between triggering and structuring causes, and between explaining something and (merely) rendering it intelligible, bring action theory closer to the philosophy of science. It is worth noting here that similar debates about the explanation of action took place in the 1960s and 1970s within the philosophy of history and social science more generally.

Responsibility and Freedom

Action theory gives rise to ethical questions concerning agency, free will, and responsibility, including the controversial *doctrine of double effect* and the *acts/omissions* distinction, both of which continue to divide consequentialists from their critics. Similar questions regarding the grounds of liability arise in criminal law.

It is a common belief that all natural phenomena are causally determined. This thought threatens our common-sense conception of ourselves as free and responsible agents. Its neuroscientific incarnation has received strong resistance from a truly diverse group of critics.

Philosophers respond to the threat of determinism in various ways, the most popular of which are "soft" compatibilist determinism, "hard" incompatibilist determinism, incompatibilist libertarianism, and agnosticism. Some theorists also find confused consolation in the thought that the randomness of indeterminism somehow leaves space for free agency. The fact that nobody adopts compatibilist nondeterminism is arguably a telling sign of the motivations behind some of the aforementioned positions.

If by "free will" we mean "free to act as one chooses or desires," there is no question that many people often act freely. But an equally legitimate and considerably more demanding understanding of free will requires its bearer to be a *source* of choice and action in a sense that would not allow for agent causation to be reducible to event causation. The possibility of such freedom is challenged by *causa sui* arguments that aim to show that such a phenomenon is epiphenomenal since it requires us to be self-causes, and not even divine beings can be such things. Such Nietzschean pronouncements may be tempered with the Sartrean thought that since not even God could act in a vacuum, the factors that make up our "facticity" are in fact *enablers* of free action.

A third notion of free will is that of having the power to do otherwise at any given moment. Harry Frankfurt has argued that such a power is not required for moral responsibility, since the latter is a function of the *reasons* we act upon. Attempts to render psychological determinism compatible with responsibility, so conceived, include appeals to whole-heartedness and second-order desires, self-reflection and satisfied plans/intentions, the semblance of responsiveness to reason, the capacity for critical evaluation, self-awareness and the desire to act in accordance with reasons, and guidance control. These models all share a commitment to an identification principle that states that we are only responsible for those beliefs and feelings that we consciously identify with, regardless of their origin, but they are divided on the issue of whether this identification is a matter of desire or normative judgment. On this issue depends the understanding of further phenomena, such as those of weakness of will and of acting against one's better judgment. The overall lesson to draw here is that agency and free will come in shades and degrees that we do not all share equally.

Constantine Sandis

See also Agency; Behaviorism in Psychological Explanation; Causes Versus Reasons in Action Explanation; Decision Theory; Determinism; Events; Explanation Versus Understanding; Free Will, Philosophical Conceptions of; Free Will in the Social Sciences

Further Readings

Aguilar, J. H., & Buckareff, A. (Eds.). (2009). *Philosophy of action: 5 questions*. Birkerød, Denmark: Automatic Press.

Anscombe, G. E. M. (1957). *Intention*. Oxford, England: Blackwell.

Austin, J. L. (1967). *How to do things with words*. Oxford, England: Clarendon Press.

Baier, A. (1971). The search for basic actions. *American Philosophical Quarterly, 8*(2), 161–170.

Bennett, J. (1988). *Events and their names.* Oxford, England: Clarendon Press.
Danto, A. C. (1973). *Analytic philosophy of action.* Cambridge, England: Cambridge University Press.
Davidson, D. (1980). *Essays on actions and events.* Oxford, England: Oxford University Press.
Dretske, F. (1988). *Explaining behaviour.* Cambridge: MIT Press.
Frankfurt, H. G. (1988). *The importance of what we care about.* Cambridge, England: Cambridge University Press.
Geach, P. (1971). *Mental acts.* London, England: Routledge & Kegan Paul.
Goldman, A. I. (1970). *A theory of human action.* Princeton, NJ: Princeton University Press.
Hornsby, J. (1980). *Actions.* London, England: Routledge & Kegan Paul.
Mele, A. R. (Ed.). (1997). *The philosophy of action.* Oxford, England: Oxford University Press.
O'Connor, T., & Sandis, C. (Eds.). (2010). *A companion to the philosophy of action.* Oxford, England: Wiley-Blackwell.
Prichard, H. A. (1949). Acting, willing, desiring. In *Moral obligation* (pp. 187–198). Oxford, England: Clarendon Press. (Original work published 1935)
Sandis, C. (Ed.). (2009). *New essays on the explanation of action.* Basingstoke, England: Palgrave Macmillan.
Sandis, C. (2012). *The things we do and why we do them.* Basingstoke, England: Palgrave Macmillan.
Thalberg, I. (1977). *Perception, emotion and action: A component approach.* Oxford, England: Blackwell.
Thompson, M. (2008). *Life and action.* Cambridge, MA: Harvard University Press.
Thomson, J. J. (1977). *Acts and other events.* Ithaca, NY: Cornell University Press.
von Wright, G. H. (1963). *Norm and action.* London, England: Routledge & Kegan Paul.
White, A. R. (Ed.). (1968). *The philosophy of action.* Oxford, England: Oxford University Press.
Wittgenstein, L. (1953). *Philosophical investigations* (G.E.M. Anscombe, Trans.). Oxford, England: Blackwell.

Actor-Network Theory

Actor-network theory (ANT) is the name given to collective Franco-British efforts at renewing social theory tout court, including its epistemology and its metaphysics. Initiated in the early 1980s by Bruno Latour and Michel Callon at the École des Mines in Paris, ANT was later extended into the United Kingdom via the efforts of John Law (and others). The theory was born through close encounters with the techno-sciences, as part of an interdisciplinary field of science and technology studies. This accounts in part for its most distinctive tenet, namely, the central role afforded material objects in the constitution of social life. ANT has since become a full social theory, promising to undo all of the latter's Great Divides: agency/structure, micro/macro, subject/object, culture/nature. As social theory, ANT seeks to reassemble the social.

In terms of theory formation, ANT is more a family resemblance of shared philosophical and methodological sensibilities than a tightly knit conceptual framework. Since its inception, the collective of ANTs has been sensitive to naming. Some, like John Law, prefer to speak of a *material semiotics* and others, like Annemarie Mol, of *ontological politics*. On his part, ANT preeminent Bruno Latour once quipped humorously that there are only four problems with ANT: the word *actor,* the word *network,* the word *theory*—and the hyphen! While agreeing that "actant–rhizome ontology" would be a more precise label, however, Latour remains ready to defend his creation, given appropriate conceptual care.

To start from one end, the actor of ANT bears little resemblance to those flesh-and-blood humans of which social theory is full. Rather, the ANT actor is always an *actant,* a semiotic entity to which action capacities are ascribed or delegated during the course of collective affairs. While I may (or indeed may not) be an actor, so may the French state, a stone, IBM, Popeye, whales, or any other *figuration,* at once semiotic and material, that interferes with the distributed action of situations and events. Moreover, all this depends on the *relations* established among entities. According to ANT, any actant attains its identity from the relations it enters into and within which it is set; this is true, as well, for human intentionality and subjectivity.

This brings us to the network. With ANT, network is neither a social network of friendship, family, or professional ties nor a technological network like the Internet or London Underground. Instead, "network" designates *any* association or assemblage of heterogeneous human and nonhuman elements. For Louis Pasteur to become the scientific icon that he is today, he needed first to enroll the microscope,

the anthrax bacillus, the vaccine, and the pest-ridden late-19th-century French agriculture into his laboratory network. In the process, the Pasteur network grew in length and size; he started macrostructuring the whole of France. Like the *rhizomes* of the French philosopher Gilles Deleuze, ANT networks are dynamic and changing, and they crisscross scales from local to global.

On to the hyphen. As Latour himself acknowledges, the hyphenated actor-network is easily mistaken as yet another take on the agency/structure debate, running deep in social theory from Émile Durkheim to Pierre Bourdieu. Just as actants are not agents, however, networks are not structures. Actor and network designate two sides of the same coin: When Pasteur acts as the *spokesperson* of his laboratory, he is himself a *black-boxed* network, the sum total of his many relations. ANT bypasses the agency/structure and micro/macro distinctions all at once. A macrostructuring actor like Pasteur is a micro-actor sitting on top of numerous more or less stable techno-scientific black boxes (microscopes, vaccines, bacilli). In ANT, power means the power of association.

Why is this not straightforwardly a theory? For ANT, the goal of social theory is not to *explain* the social by invoking structures, fields, or systems; instead, the goal is to *describe* how the social is gradually composed and reassembled. Society explains nothing; it must itself be explained. Such is the revenge, Latour asserts, of Gabriel Tarde over Émile Durkheim a hundred years on. To see how the social is reassembled, one must *follow the actors,* not impose on them a preset metatheory. Like the ethnomethodology of Harold Garfinkel, ANT seeks to learn from the ethnomethods of social actors. Rather than a theory, it is a method for deploying and mapping the world-building capacities of actors. With ANT, actors contain their own variable ontology, even their own metaphysics; this is why Latour describes his aim and style as *empirical philosophy.*

On this note, Latour plays a philosophical double game. While his ANT invites actors to define the social ontology for him, he clearly has no hesitation in engaging in serious philosophizing himself. Most famously, Latour (1993) takes his studies in the anthropology and sociology of science as evidence that we, the West, have never been modern. Modernity, to Latour, consists in the ontological dictum of keeping nature separate from culture, facts from values, science from politics. The many *hybrids* generated in techno-scientific laboratories, however, have never respected such boundaries; in a world of ecological crises—of ozone holes, nuclear radiation, and climate change—such nature–culture infiltrations are increasingly obvious. In the end, ANT is the social theory needed to trace the many hybrids of our *non-modern* (not post-modern) world.

In terms of its ability to form a heterogeneous academic network, ANT represents a social science success story. Concepts from ANT have been taken up widely across a range of disciplines, and its methods have been brought to bear on an expanding range of topics. Well beyond the world of science and technology, there are now actor-network–inspired takes on everything from art making to lawmaking, from organization studies to religious studies.

In this process of translation, ANT encounters a range of obstacles, as social theorists and philosophers object to its post-humanism, its relativism, or its lack of critical distance. Latour, however, rejects these terms of debate: ANT is *nonhumanist,* not post-humanist; it is *relationalist,* not relativist; and it seeks *critical proximity,* not critical distance. Ultimately, ANT thus demands nothing less than a new social-philosophical vocabulary, as it embarks on the adventure of non-modernity.

Anders Blok

See also Contemporary French Philosophy and the Social Sciences; Ethnomethodology; Science and Ideology; Social Objects Versus Technical Objects; Social Studies of Science and Technology; Technoscience and Society

Further Readings

Callon, M. (1991). Techno-economic networks and irreversibility. In J. Law (Ed.), *A sociology of monsters: Essays on power, technology and domination* (pp. 132–165). London, England: Routledge.

Latour, B. (1993). *We have never been modern.* Cambridge, MA: Harvard University Press.

Latour, B. (1999). On recalling ANT. In J. Law & J. Hassard (Eds.), *Actor network theory and after* (pp. 15–25). Oxford, England: Blackwell.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network theory.* Oxford, England: Oxford University Press.

Law, J. (2008). Actor-network theory and material semiotics. In B. S. Turner (Ed.), *The new Blackwell companion to social theory* (3rd ed., pp. 141–158). Oxford, England: Blackwell.

Mol, A. (2010). Actor-network theory: Sensitive terms and enduring tensions. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 50*(1), 253–269.

Affective Intelligence in the Social Sciences

Affective intelligence refers to a set of skills and abilities that allows individuals to recognize, manage, respond to, and communicate *emotions* effectively. There has been growing popular recognition of the pivotal role of affect in social life, matched by extensive interest in affectivity in the social sciences, including economics, sociology, political science, and education. Psychological research aimed at measuring and defining this construct is a relatively recent development.

This entry introduces the novel field of affective intelligence, explains the notion, points to its historical precursors, and shows the importance of affect and emotions in affective intelligence, social cognition, and social-scientific inquiry in general.

Within Western philosophy, philosophers such as Aristotle, Socrates, Plato, St. Augustine, the Stoics, René Descartes, Blaise Pascal, Baruch Spinoza, Adam Smith, and Immanuel Kant have all sought to understand the role of affect in general and emotions in particular in thinking and behavior. Plato, for example, thought that affect constitutes a more primitive, animal aspect of human nature that is incompatible with reason and that the latter should direct or subjugate the former. The idea that affect subverts rational thinking was perpetuated in the speculative ideas of Sigmund Freud and others. Some writers, such as Arthur Koestler, even thought that humans' inability to know and control their emotions suggests a fatal "flaw" in the way their brains developed, an evolutionary mistake that may threaten the very survival of our species. Surprisingly, most of what we know about the role of affect in social cognition and behavior has only been discovered recently. Although feeling and thinking were often assumed to be separate faculties by early philosophers and psychologists, recent research suggests a fundamental interdependence between affect, cognition, and behavior.

In the past few decades, neuroscientists and psychologists produced incontrovertible evidence that affect is not only not dangerous but does in fact provide essential and adaptive information necessary for dealing with the manifold challenges of social life. For most of the history of the social sciences, *intelligence* and *affect* were considered as completely unrelated areas. The development of empirical tests of intelligence brought rapid progress in that field, but research on affect remained neglected during the dominance of behaviorist and cognitivist paradigms in psychology. Charles Darwin's view of emotions as biologically determined and serving an adaptive purpose was not really taken seriously until the 1970s by evolutionary sociologists and psychologists. By the 1980s, many researchers on human cognition came to embrace the notion that affect is a necessary and integral part of all thinking, and there was a rapid explosion in research exploring the interaction between affect and cognition. Cognitive scientists showed growing interest in incorporating affective reactions into their artificial intelligence models, and there was also renewed interest in studying the evolutionary, adaptive functions of emotions.

In a sense, the concept of affective intelligence emerged before there was a proper empirical approach to define and measure the construct. It was not until the 1990s that psychologists such as Peter Salovey and Jack Mayer proposed a theory and *measurement* of affective intelligence, integrating work from neuroscience, psychology, philosophy, and clinical psychology. Popular writers such as Daniel Goleman soon claimed that affective intelligence is one of the most important predictors of personal and professional success in life, although its nature and characteristics remained fuzzy and poorly defined. Despite impressive research efforts to measure and define affective intelligence, there remains a fundamental disconnection between the notion of affective intelligence as used in the popular media and the available scientific evidence supporting this concept. There have been a variety of conflicting definitions and empirical approaches, and there is still no reliable consensus on what the term means and how it should be measured.

For example, one might define and measure affective intelligence as a number of distinct

affective competencies assessable in an inventory, although it remains unclear if these competencies are related and necessarily occur together in the same person. Others define affective intelligence as encompassing all *noncognitive* abilities or skills that influence one's ability to succeed in coping with environmental demands and pressures. This approach resulted in the measurement of a variety of often unrelated personal characteristics, such as empathy, flexibility, happiness, impulse control, self-regard, and the like.

Alternatively, affective intelligence has been defined by Mayer and Salovey as a general, underlying ability to recognize the meanings of emotion and their relationships and to reason and problem solve on the basis of them. Affective intelligence is probably involved in the capacity to perceive emotions, assimilate emotion-related feelings, understand the information of those emotions, and manage them. Mayer and Salovey's test of affective intelligence measures four kinds of abilities: the ability to (1) perceive emotion, (2) use emotion to inform thinking, (3) understand emotional meanings, and (4) manage emotions in oneself and others.

It is still unclear if affective intelligence refers to a single, coherent, and definable individual characteristic or whether it is a term that encompasses a variety of already known individual abilities that simply have not been considered in conjunction before. There is some evidence that some affective intelligence scales do measure characteristics that are different from those measured by either existing intelligence scales or personality measures, or are not well captured by them. However, some of the other scales and definitions of affective intelligence appear highly correlated with existing personality characteristics and thus turn out to be redundant.

Another critical question concerns the usefulness of affective intelligence in predicting important future outcomes, as does, for example, intelligence. Daniel Goleman even suggested that affective intelligence may be more important than the intelligence quotient (IQ) in predicting everyday success. Unfortunately, the evidence has failed to support such claims. Some empirically minded psychologists came to despair of the concept and concluded that little remains of emotional intelligence that is unique and psychometrically sound.

Despite the fact that affective intelligence is undoubtedly a complex, multifaceted, and hard-to-measure construct, and is not particularly effective in predicting future outcomes, it nevertheless remains an important concept in the social sciences in general and psychology in particular. Understanding how affect functions in informing us about the demands we face in everyday situations remains a critically important question in economics, education, psychology, and philosophy. George Marcus and others have suggested that affective intelligence is also a key concept in understanding *political judgments* and behavior, such as voting and party identification. Furthermore, studying affective intelligence can be a useful avenue for improving self-knowledge and our understanding of others; it should help people recognize and act upon emotional signals; and it can contribute to improved awareness of the importance of emotions in organizational, clinical, educational, and interpersonal situations.

Joseph P. Forgas

See also Artificial Intelligence; Cognitive Sciences; Emotions; Emotions in Economic Behavior; Intelligence; Social Cognition; Social Neuroscience

Further Readings

Ciarrochi, J., Forgas, J. P., & Mayer, J. D. (Eds.). (2006). *Emotional intelligence in everyday life* (2nd ed.). New York, NY: Psychology Press.

Forgas, J. (Ed.). (2001). *Handbook of affect and social cognition.* Mahwah: NJ: Lawrence Erlbaum.

Goleman, D. (1998). *Working with emotional intelligence.* New York, NY: Bantam.

Marcus, G. E, Neuman, W. R., & MacKuen, M. (2000). *Affective intelligence and political judgment.* Chicago, IL: University of Chicago Press.

Matthews, G., Zeidner, M., & Roberts, R. D. (2004). *Emotional intelligence: Science and myth*. Cambridge: MIT Press.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry, 60,* 197–221.

Agency

Agency is the property of being an agent. An agent is an entity that acts. This makes acting or action the primary notion in terms of which agency is defined.

Actions and Agents

There are more and less inclusive uses of the word *action* in ordinary English. In a very inclusive use, there are the actions of acids, waves, and winds. For example, acids dissolve things, waves push and drag things, and winds blow things around. Events of these kinds count as actions in a broad sense. Philosophers tend to have little interest in events that are actions only in this very broad sense. They are primarily interested in the sphere of intentional actions. In this connection, they are also interested in unintentional actions: Part of what is required for understanding what intentional actions are is understanding how they differ from unintentional actions.

Agents may be divided into kinds along two different dimensions. On the one hand, there are human agents, canine agents, feline agents, and so on. Here, the division is biological. (If chemical agents are included, the principle of division is broader.) On the other hand, there may be agents that are capable of exercising free will and agents that are not, agents that are capable of constructing and executing plans for the distant future and agents that are not, agents that are capable of acting intentionally and agents that are not, and so forth. Here, the division is based on agential capacities. An interesting question that has not received much philosophical attention is how agency of the second kind is related to agency of the first kind. For example, how far does the capacity for intentional action extend in the animal kingdom?

Intentional Action and Intentions

How the question just raised about intentional action in the animal kingdom is to be answered depends on what intentional action is. It is plausible that only agents that have intentions are capable of acting intentionally. So how far do intentions extend in the animal kingdom? That depends, of course, on what intentions are.

Intentions have been a topic of considerable discussion not only in philosophy but also in neuroscience, various branches of psychology, and legal theory. It is fair to say that no single definition of the term is widely used. According to one view of intentions, their contents are plans and one of their functions is to contribute to the execution of these plans. The plans are representations of what is intended.

Intentions may be divided into different kinds. One main division is between proximal and distal intentions—intentions about what to do now and intentions about what to do later. Many human beings have distal intentions that aim years into the future. A student's intention to graduate four years from now is a case in point. How far into the future intentions of nonhuman animals can aim is a topic for investigation.

Intending to do something should be distinguished from merely wanting to do it. Ann wants to have dinner with Bob today at 6 p.m., as she always does after their Tuesday seminar; but she also wants to hear a lecture that begins at 6 p.m. She is undecided about what to do. She can settle matters by forming a relevant intention—for example, an intention to attend the lecture. That is what she does.

In Ann's case, competing desires generate uncertainty about what she should do, which she resolves by forming an intention to attend the lecture. At what point in their development do human children start resolving uncertainty generated by competing desires by forming intentions? Which kinds of animal do this? These are topics for investigation.

Human Agency

Understanding human agency is a partly conceptual and partly empirical enterprise. Important questions include, but are not limited to, the following. How are intentional human actions produced? Does anyone have free will? Are we ever morally responsible for what we do? Given persisting disagreement about the meaning of some of the crucial terms, it may be thought that we are in no position to conduct scientific investigations of some of these topics. But this is unduly pessimistic.

Take free will as an example. According to some ways of understanding it, an agent whose conscious intentions are never among the causes of corresponding actions lacks free will. When this is assumed about free will, scientific investigations of connections—or the absence thereof—between conscious intentions and the corresponding actions are relevant to free will. And there are such investigations. In his book *The Illusion of Conscious Will,* Daniel Wegner argues, on the basis of various data, that conscious intentions are never among the causes of corresponding actions. On the other side of the issue, studies of implementation intentions provide evidence that conscious intentions (or their neural correlates) sometimes are among the causes of

corresponding behavior. In the case of subjects who share a certain distal goal—for example, conducting a breast self-examination the following month or exercising vigorously for 20 consecutive minutes the next week—those who are instructed to decide now on a place and time to pursue that goal later have a much higher success rate than those who are not given this instruction. In the aggregate, the conscious implementation decisions or intentions seem to make a significant behavioral difference.

The production of intentional human actions is another good example of something that can be investigated scientifically, even though there is significant disagreement about the meaning of a key term—*intentional action,* in this case. There is an ancient philosophical theory, advanced in Plato's *Protagoras* and elsewhere, according to which it is not possible for human beings to willingly (or freely) act contrary to what they know is best. According to this theory, when people seem to be doing this, either they are acting unwillingly (think of a compulsive hand washer who washes his hands even though he is convinced that he should not be doing that again so soon) or they have come to believe that what they are doing is best. Hard evidence about how intentional human actions are produced may be expected to shed light on whether this theory should be accepted or rejected.

There is evidence that the motivational strength or action-causing power of our desires is influenced not only by our beliefs about what is good but also by factors such as the perceived proximity of prospects for desire satisfaction, the salience of the desired objects in perception or in imagination, and the way we attend to the desired objects. The evidence indicates that what we are most strongly motivated to do at a time may be out of line with what we judge it best to do then, even when we are free from compulsion. In such a case, if we act as we are most strongly motivated to act, we falsify the ancient theory.

Perhaps in some cases in which we are tempted to act contrary to what we deem best on the whole, we can go either way. A full-blown theory about how human actions are produced will tell us something about how actions are produced in these situations. Normal agents apparently can influence the strength of their desires in a wide variety of ways. For example, they can refuse to focus their attention on the attractive aspects of a tempting course of action and concentrate instead on what is to be accomplished by acting as they judge best. They can attempt to augment their motivation for performing the action judged best by promising themselves rewards for doing so. They can picture a desired item as something unattractive—for example, a wedge of chocolate pie as a wedge of chewing tobacco—or as something that simply is not arousing. Desires typically do not have immutable strengths, and the plasticity of motivational strength is presupposed by standard conceptions of self-control. Intentional self-intervention—action aimed at changing one's own motivational condition—is a topic for further philosophical and scientific investigation.

Self-Deception

Ordinary human agents seem to have a capacity for self-deception. Predictably, philosophers disagree about what "self-deception" means. But the competing accounts of its meaning have been set out with some precision, and we can look for hard evidence on whether or how self-deception happens, given the various accounts. One view about what self-deception means is inspired by Sigmund Freud: At the view's core is the idea that a person who is self-deceived regarding some proposition *p* subconsciously believes that *p* is false while consciously believing that *p* is true. We can look for hard evidence on whether this ever happens. According to an alternative view (one of several alternatives), self-deception is, very roughly, motivationally or emotionally biased false belief. Obviously, on this view, scientific work on the influence of motivation and emotion on belief is relevant to understanding how and why self-deception happens.

A related disagreement about what self-deception is (or means) features intentional action. In garden-variety cases of self-deception, do people *intentionally* deceive themselves? Proponents of agency views of self-deception answer yes; their opponents—anti-agency theorists—answer no. One can imagine relatively clear cases of intentionally deceiving oneself into believing that *p* is true. A student who made a fool of himself in class today but would like to believe that he performed well writes the following in his diary: "I was awesome in math class today!" He knows both that he has a bad memory and that his policy is to read his entries for a semester after that semester ends. A month later, he reads his diary entry for the day at issue and believes what he wrote,

having forgotten why he wrote it. But such cases seem remote from garden-variety self-deception.

Disputes About Meaning and Experimental Philosophy of Action

When philosophers disagree about the meanings of terms, they sometimes claim that their opponents have strayed unacceptably far from ordinary usage. Experimental philosophy of action is a relatively new development that aims to test such claims; to shed light on ordinary usage of terms such as *action, intentional action,* and *free will*; and to gather evidence about what influences people's judgments about things such as whether an agent did something intentionally or had free will when he did it. The method features presenting nonspecialists with vignettes and asking questions about the agents in them—for example, whether a character in a vignette intentionally hit a target or deserved to be blamed for some harm she caused. Much of experimental philosophy of action is done by interdisciplinary teams of researchers—especially philosophers and psychologists.

Group Agency

Agency is an enormous topic, even when it is limited to individual agents, as has thus far been the case here. But we talk as though agency extends to groups—for example, teams and corporations. In this sphere too, acting or action is the basic notion. Is a team an agent? If so, that is because a team is an entity that acts. How actions of groups, if there are any such actions, are related to actions of individuals is an intriguing question that lies beyond the scope of this entry.

Alfred Mele

See also Action, Philosophical Theory of; Collective Agents; Experimental Philosophy; Free Will, Philosophical Conceptions of; Intention, Social Psychology of; Unconscious; Unconscious Social Behavior

Further Readings

Baumeister, R. (2002). Ego depletion and self-control failure: An energy model of the self's executive function. *Self and Identity, 1,* 129–136.

Gollwitzer, P. (1999). Implementation intentions. *American Psychologist, 54,* 493–503.

Knobe, J., & Nichols, S. (Eds.). (2008). *Experimental philosophy.* Oxford, England: Oxford University Press.

Mele, A. (2001). *Self-deception unmasked.* Princeton, NJ: Princeton University Press.

Mele, A. (2003). *Motivation and agency.* New York, NY: Oxford University Press.

Mischel, W., & Ayduk, O. (2004). Willpower in a cognitive-affective processing system. In R. Baumeister & K. Vohs (Eds.), *Handbook of self-regulation* (pp. 99–129). New York, NY: Guilford Press.

Plato. (1976). *Protagoras* (C. Taylor, Trans.). Oxford, England: Clarendon Press.

Wegner, D. (2002). *The illusion of conscious will.* Cambridge: MIT Press.

Agent-Based Modeling and Simulation in the Social Sciences

This entry reviews agent-based modeling (ABM) with respect to three aspects: (1) its main features and relationship to other simulation techniques, (2) the distinctiveness of the ABM methodology in terms of its approach to the problems of agency and explanation, and (3) open questions for practitioners in the field. The entry ends with a list of useful resources available on the web.

The Field of ABM

ABM is a methodology for the study of social phenomena by exploring, through the use of computational models, the relations connecting microbehavior with macro patterns. This methodology is increasing in popularity for three reasons. The first is the increasing technical advances in computing capability since the last decades of the 20th century. The second is the flexibility of programming languages compared with mathematical and logical formalisms. The third is the growing interest in features such as complexity, emergence, self-organization, nonlinearity, and path dependence in the social sciences.

ABM is framed in a wider approach to social phenomena: computational social simulation. Computational simulation allows one to avoid both the vagueness and ambiguity of verbal description and the stringent assumptions of purely mathematical description. It works through the design of a model intended to mimic or replicate the desired

social phenomenon in a particular way. The model can then be subjected to experimentation by the manipulation of prespecified parameters.

Computational models to study social phenomena can be divided into three types: macro-simulation, micro-simulation, and agent-based simulation. In the first, models are used to explore system dynamics, a name often used to describe this type of simulation. Such models use sets of difference or differential equations to estimate the behavior of system variables over time. The *macro* prefix is used because this estimation is performed from a top-down perspective: The target system is simulated as a whole. Individuals are not taken into account in the model but only population-level attributes. These attributes are thought of as causally connected, and causal relations are represented in the model's equations. During the simulation, the equations calculate the values for each variable at every time step, according to the previous value of the variables with causal influence.

In micro-simulation, the estimation also depends on sets of equations, but the opposite approach, bottom-up, is employed. This means that estimation is based not on the values of system variables but on the attributes of the basic units (individuals, households, organizations, etc.). Micro-simulation works by applying transition rules that change these individual attributes as time passes. The individual results are then aggregated into a hypothetical sample that represents the overall population once the simulation is over.

Although different in their approach, micro- and macro-simulations are both simulation techniques in which models are fed with empirical data and in which the main concern is prediction. ABM is grounded in a different principle. As with micro-simulation, ABM uses a bottom-up approach. However, in ABM, the basic units have the possibility to interact. This important feature has channeled the discussion about ABM into two specific topics: (1) the characterization and role of the basic units—that is, the agents in the models—and (2) the way in which the tension between explanation and prediction in social science is approached with ABM.

Agency and Explanation

The notion of "agent" in ABM is less elaborated than is typical in philosophy. In the latter, agency is usually associated with intricate issues such as free will, determinism, and so on. In ABM, "agent" simply refers to a self-contained program—in other words, an independent portion of computer code with the ability to choose an action according to its perception of the surrounding environment. This "intelligent" behavior is possible because agents can perform symbolic manipulation according to some coded set of instructions. There are three ways to build agents that can do this: (1) direct programming, using the features of object-oriented programming languages to code agents and the environment; (2) production systems, in which agents are created as structures composed by a set of rules, a working memory, and a rule interpreter; or (3) neural networks, mimicking the operation of a nervous system, to simulate agents as the nodes of neural networks or as neural networks themselves.

Although there are some differences in the way they work, all three methods allow for the construction of an agent that can perceive the environment, record information about it, and generate behaviors (e.g., communication, movement, and action on the environment). This feature makes it possible to use ABM to analyze situations where there is no central coordination in populations of heterogeneous agents. In these models, social properties or patterns are studied as emerging from the interactions of individuals with each other and with the environment.

ABM tends to emphasize explanation above prediction. Positivistic philosophy gave a decisive role to prediction, which was commonly equated with explanation. Post-positivistic stances criticized this view and called for more comprehension-based explanations. ABM is a suitable methodology for comprehension for it allows for focusing on the "how," by representing entities directly in interactions and by simulating one process with another. There is an ontological correspondence between agents in the model and agents in real life. The same is true with the environment when it is spatially explicit—that is, with a direct representation of the objects surrounding the agents.

ABM is also oriented to the "why"—that is, etiological explanation—explaining how a phenomenon is brought about, that is, referring to its cause(s). The most common approach to explanation in ABM is the *generative* approach. According to this, to explain a phenomenon one has to "grow" it. This means recreating the initial conditions at the micro level and running the simulation in order to check if, through interaction, the desired macro property or pattern is obtained. This approach has found theoretical

support in the mechanist perspective in the social sciences, which is also focused on the analysis of entities and interactions through time. There are some issues still to be clarified in terms of compatibility, but this theoretical framework could make explanation through simulation more robust in the future.

Critical Questions

In addition to the adoption of the mechanist perspective, there are other issues that remain as open questions. ABM is usually considered as combining some of the features of qualitative and quantitative methodologies. Some argue that simulation could also be used as a tool for developing formal theory. Formal theory is usually seen as a way to avoid ambiguity and vagueness both in concepts and in causal links. In the social sciences, however, formalization is quite uncommon, except in some areas of economics. ABM could be used for formalization for four reasons. Compared with mathematical equations or systems, (1) programming languages are syntactically and semantically richer and (2) allow for the construction of independently separated segments of code that can be easily altered or removed, in order to change the behavior of the model. This flexibility also allows for (3) the design of heterogeneous agents that (4) can act in parallel in the environment, with no specific sequential order. This theoretically driven conception of ABM, despite its possible advantages, is rarely explored, not only because of the dominance of the methodologically driven view but also because of the lack of theoretical foundations for the current approach. As the value of simulation becomes better appreciated and more social scientists get involved with modeling, further advances could be achieved in this regard.

Although ABM was thought of initially as a tool for theoretical exploration with few predictive ambitions, some researchers in the field have argued in favor of more complex models, based heavily on real data to calibrate the simulation. This shift has an epistemological basis, for some believe that traditional models are too simple to yield useful information, but it is also being advocated for practical reasons, because ABM is now being used in areas where more complex models might be desirable, such as public policy. The debate has been labeled as KISS (Keep It Simple, Stupid) versus KIDS (Keep It Descriptive, Stupid). In the former camp, models with few parameters and agents with low cognitive capabilities still dominate. It is believed that the more complex the model, the harder it is to identify the mechanisms linking microbehaviors and macro patterns. In the latter camp, more realistic interaction rules and agent behavior are considered desirable. Some researchers, however, develop strategies that move between these two poles and link theory and empirical data in two different ways. In some cases, simple theoretical models are generated and then tested against empirical data; in others, empirically calibrated models are built, and then theoretical insights are derived from them.

This debate is linked to a third open discussion on ABM: verification and validation. The flexibility in the way computer models are built has led to an accumulation of models that, although dealing with the same topic, have parameter specifications that make them incommensurable. This is even more problematic in the cases where model results are significantly path sensitive. When the model is calibrated with real data and the model is intended to replicate observable social phenomena, verification and validation become easier, for the output can be compared with the empirical data. Empirical cross-validation, however, does not work in every case. ABM does not assume the linearity of the problem analyzed, which can be another source of incommensurability if results are compared with those produced by the more traditional analytical methods that rely on the general linear model.

There has been an increasing concern with internal and external methods of validation, including not only scientific but also social criteria for verification and validation. Model verification and validation is now seen more as a process in which interaction and feedback from different stakeholders can provide valuable insights in terms of consistency and accuracy, not simply as a close scientific matter that researchers can solve by themselves with a prescribed set of evaluations and procedures. Still, this remains an open topic for discussion inside the simulation community and has been a thorny issue in terms of the acknowledgment of ABM as a proper scientific tool for inquiry.

Resources

Several free online resources are available about social simulation. One option is to download and try

demonstration models using a simulation toolkit. The three most popular are Netlogo (http://ccl.northwestern.edu/netlogo), Repast (http://repast.sourceforge.net), and Mason (http://cs.gmu.edu/~eclab/projects/mason). They each include resources such as libraries, sample models, and documentation. Other websites, such as OpenABM (http://www.openabm.org), Agent-Based Computational Economics (http://www.econ.iastate.edu/tesfatsi/ace.htm), and Agents-Based Models: Methodology and Philosophy (http://www.agent-based-models.com/blog), contain useful information about software and about researchers and institutions working with ABM. There are two useful e-mailing lists that regularly feature information about events, publications, and jobs: SIMSOC (http://www.jiscmail.ac.uk/lists/simsoc.html) and Complexity Digest (http://www.comdig.org).

Nigel Gilbert and David Anzola

See also Agency; Analytical Sociology and Social Mechanisms; Artificial Intelligence; Complexity; Explanation, Theories of; Game-Theoretic Modeling; Mechanism and Mechanismic Explanation; Microfoundationalism; Models in Social Science; Multi-Agent Modeling

Further Readings

Epstein, J., & Axtell, R. (1996). *Growing artificial societies*. Cambridge: MIT Press.

Gilbert, N. (2007). *Agent-based models*. London, England: Sage.

Gilbert, N. (Ed.). (2010). *Computational social science*. Thousand Oaks, CA: Sage.

Gilbert, N., & Troitzsch, K. (2005). *Simulation for the social scientist*. Glasgow, England: Open University Press.

Miller, J., & Page, S. (2007). *Complex adaptive systems*. Princeton, NJ: Princeton University Press.

Agnotology, Ignorance, and Uncertainty

Agnotology is the study of ignorance (from the Greek word *agnoia* (ἄγνοια), meaning "ignorance"; *a* = "non" and *gnosis* = "knowledge"). The term is a neologism coined by Robert Proctor.

Ignorance, uncertainty, and related terms refer variously to the absence of knowledge, doubt, and false belief. This topic has a long history in Western philosophy, famously rooted in the Socratic tradition. It has a considerably shorter and, until recently, sporadic treatment in the human sciences. The focus of this entry is on relatively recent developments within and exchanges between both domains.

A key starting point regarding ignorance is that anyone attributing it cannot avoid making claims to know something about who is ignorant of what: A is ignorant from B's viewpoint if A fails to agree with or show awareness of ideas that B defines as actually or potentially valid. A and B can be identical, so that A self-attributes ignorance. Numerous scholars thereby have noted the distinction between conscious ignorance (known unknowns, learned ignorance) and meta-ignorance (unknown unknowns, ignorance squared).

The topic has been beset with terminological difficulties due to the scarcity and negative cast of terms for referring to unknowns. Several scholars have constructed typologies of unknowns, in attempts to make explicit their most important properties. Smithson's book, *Ignorance and Uncertainty: Emerging Paradigms,* pointed out the distinction between being ignorant of something and ignoring something, the latter being akin to treating something as irrelevant or taboo. Karin Knorr-Cetina coined the term *negative knowledge* to describe knowledge about the limits of the knowable. Various authors have tried to distinguish reducible from irreducible unknowns.

Two fundamental concerns have been at the forefront of philosophical and social-scientific approaches to unknowns. The first of these are judgment, learning, and decision making in the absence of complete information. Prescriptive frameworks advise how this ought to be done, and descriptive frameworks describe how humans (or other species) do so. A dominant prescriptive framework since the second half of the 20th century is subjective expected utility theory (SEU), whose central tenet is that decisional outcomes are to be evaluated by their expected utility; that is, the product of their probability and their utility (e.g., monetary value, although utility may be based on subjective appraisals). According to SEU, a rational decision maker chooses the option that maximizes her/his expected utility. Several descriptive theories in psychology and behavioral economics (e.g., Prospect Theory

and Rank-Dependent Expected Utility Theory) have amended SEU to render it more descriptively accurate while retaining some of its "rational" properties.

The second concern is the nature and genesis of unknowns. While many scholars have treated unknowns as arising from limits to human experience and cognitive capacity, increasing attention has been paid recently to the thesis that unknowns are socially constructed, many of them intentionally so. Smithson's 1989 book was among the earliest to take up the thesis that unknowns are socially constructed. Related work includes Robert Proctor's 1995 *Cancer Wars* and Ulrich Beck's 1992 *Risk Society.* Early in the 21st century, this thesis has become more mainstream. Indeed, a 2008 edited volume, bearing *agnotology* in its title, focuses on how culture, politics, and social dynamics shape what people do not know.

Philosophers and social scientists alike have debated whether there are different kinds of unknowns. This issue is important because if there is only one kind then only one prescriptive decisional framework is necessary and it also may be the case that humans have evolved one dominant way of making decisions with unknowns. On the other hand, different kinds of unknowns may require distinct methods for dealing with them.

In philosophy and mathematics, the dominant formal framework for dealing with unknowns has been one or other theory of probability. However, Max Black's groundbreaking 1937 paper proposed that vagueness and ambiguity are distinguishable from each other, from probability, and also from what he called "generality." The 1960s and 1970s saw a proliferation of mathematical and philosophical frameworks purporting to encompass nonprobabilistic unknowns, such as fuzzy set theory, rough sets, fuzzy logic, belief functions, and imprecise probabilities. Debates have continued to this day over whether any of these alternatives are necessary, whether all unknowns can be reduced to some form of probability, and whether there are rational accounts of how to deal with nonprobabilistic unknowns. The chief contenders currently include generalized probability frameworks (including imprecise probabilities, credal sets, and belief functions), robust Bayesian techniques, and hybrid fuzzy logic techniques.

In the social sciences during the early 1920s, John Maynard Keynes distinguished between evidentiary "strength" and "weight," while Frank H. Knight similarly separated "risk" (where probabilities are known precisely) from "uncertainty" (where probabilities are not known). Ellsberg's classic 1961 experiments demonstrated that people's choices can be influenced by how imprecisely probabilities are known (i.e., "ambiguity"), and his results have been replicated and extended by numerous studies. Smithson's 1989 book proposed a taxonomy of unknowns, and his 1999 experiments showed that choices also are influenced by uncertainty arising from conflict (i.e., disagreeing evidence from equally credible sources); those results also have been replicated.

More recent empirical research on how humans process unknowns has utilized brain imaging methods. Several studies have suggested that Knightian uncertainty (ambiguity) and risk differentially activate the ventral systems that evaluate potential rewards (the so-called reward center) and the prefrontal and parietal regions, with the latter two becoming more active under ambiguity. Other kinds of unknowns are yet to be widely studied in this fashion, but research on them is emerging. Nevertheless, the evidence thus far suggests that the human brain treats unknowns as if there are different kinds.

Finally, there are continuing debates on whether different kinds of unknowns should be incorporated in prescriptive decision-making frameworks and, if so, how a rational agent should deal with them. There are several decisional frameworks incorporating ambiguity or imprecision, some of which date back to the mid 20th century, and recently at least one incorporating conflict as well. The most common recommendation for decision making under ambiguity amounts to a type of worst-case analysis. For instance, given a lower and upper estimate of the probability of Event E, the usual advice is to use the lower probability for evaluating bets on E occurring but to use the upper probability for bets against E. However, the general issue of what constitutes rational behavior under nonprobabilistic uncertainties such as ambiguity, fuzziness, or conflict remains unresolved.

Michael Smithson

See also Abduction and Inference to the Best Explanation; Induction and Confirmation; Probability; Rational Expectations; Risk

Further Readings

Bammer, G., & Smithson, M. (Eds.). (2008). *Uncertainty and risk: Multidisciplinary perspectives.* London, England: Earthscan.

Beck, U. (1992). *Risk society: Towards a new modernity.* London, England: Sage.
Beck, U. (1999). *World risk society.* Oxford, England: Polity Press.
Black, M. (1937). Vagueness: An exercise in logical analysis. *Philosophy of Science, 4,* 427–455.
Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics, 75,* 643–669.
Gardenfors, P., & Sahlin, N.-E. (Eds.). (1988). *Decision, probability, and utility: Selected readings.* Cambridge, England: Cambridge University Press.
Keynes, J. M. (1921). *A treatise on probability.* London, England: Macmillan.
Knight, F. H. (1921). *Risk, uncertainty, and profit* (Hart, Schaffner, and Marx Prize Essays, No. 31). Boston, MA: Houghton Mifflin.
Knorr, C. K. (1999). *Epistemic cultures: How the sciences make knowledge.* Cambridge, MA: Harvard University Press.
Proctor, R. N. (1995). *Cancer wars: How politics shapes what we know and don't know about cancer.* New York, NY: Basic Books.
Proctor, R. N., & Schiebinger, L. (Eds.). (2008). *Agnotology: The making and unmaking of ignorance.* Stanford, CA: Stanford University Press.
Smithson, M. (1989). *Ignorance and uncertainty: Emerging paradigms* (Cognitive Science Series). New York, NY: Springer.
Smithson, M. (1999). Conflict aversion: Preference for ambiguity vs. conflict in sources and evidence. *Organizational Behavior and Human Decision Processes, 79,* 179–198.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297–323.
Walley, P. (1991). *Statistical reasoning with imprecise probabilities.* London, England: Chapman Hall.

Alienation: From Philosophy to Sociology

Although occasionally disparaged as a vague term, the concept of alienation has been dealt with widely in philosophy and is still important in fields as disparate as law, medicine, religion, economics, sociology, political science, psychology, as well as literature. Historically, from *alienus,* meaning "other," Roman law used "alienation" to signal that the seller has relinquished all claims to some given property. Contemporary legal scholars examine alienation as a phenomenon of administrative law, a blot on relations between administrators and the public. From medieval to early-modern times, insanity was identified as alienation from reason, with psychiatrists—"alienists"—having the power to prescribe appropriate cures or sequestration of the irrational. Theologians have dealt with alienation as a self-defeating separation from the divine. In positivist economics, alienation is important as a measure of social distance between identifiable groups of agents. Political research examines alienation in relation to the political behaviors. In literature, individuals' alienation is explored as the affective formation of personality. In sociology and in social psychology, alienation has been used to describe the operations of social structures under capitalism, as well as modes of social engagement among peers, in families, in organizational life, and in politics. All these approaches in the social sciences and the humanities stem from views of alienation as a philosophical issue.

Philosophy

Alienation may be traced as a concept to the Hebrew and Christian Bible, and in Roman thought to Plotinus's doctrine of emanation; it is significant for understanding the split between idealism and materialism, between G. W. F. Hegel and Karl Marx, and for tracing developments in social and political philosophy.

G. W. F. Hegel

In Hegel's writings, as Spirit moved through history, two terms, *Entäusserung* ("externalization") and *Entfremdung* ("estrangement") referred to alienation. Alienation as *Entäusserung* has a neutral or even a positive connotation, as it is intrinsic to the externalization of thought. Interpreted as the self-objectification of Spirit, *Entäusserung* was the basis for the dialectical contradictions whose negations, new syntheses, and subsequent negations (of negations) moved Spirit through history toward its telos—often a "journey of despair." But when understood by philosophers who seek precision and clarity of the world, absolute knowledge provided a joyous or spiritual consciousness. With absolute knowledge, alienation is overcome, since knowledge now knows itself. As Plato and Aristotle said, a life devoted to knowledge for its own sake is the "happiest" life.

Hegel's analysis of "the alienated spirit" heralded several developments. As a critic of religion, Ludwig Feuerbach (1854/1989) saw "alienation produced when the self . . . makes its own essential nature another objectified being [God]" (that stood without and acted as an alien force. For Hegel's phenomenology, the transformation of nature through the labor of the slave provided the initial moment of self-awareness as differentiation from nature and, in turn, recognition of the master. Hegel's description of the master's appropriation of the labor of the slave, and the slave's subjection of self to the master, set the stage for Marx's critique of capitalism as the bourgeois appropriation of proletarian labor. *Entfremdung,* an outcome of increasing machine mediation between humans and nature, came to have a pejorative sense.

Karl Marx

Although Hegelian idealism was the dominant philosophy of the era, Marx's interest in materialism in early Greek philosophy, as well as in the contemplative materialism of Feuerbach, led to his rejection of idealism. Still, he remained indebted to the idealists' grasp of the active side of history. Embracing the Hegelian dialectic, with its critique of domination as embodied in the Master–Slave struggle for recognition, and the central role of alienation as objectification and estrangement, Marx developed a materialist epistemology and ontology informed by a view of human beings as destined to work and refashion nature. Observing the adverse effects of capitalist modernity, Marx inverted Hegel's idealism in favor of a materialist basis for the subject as well as ideas and moved alienation from its ontological status into a sociological moment as the basis for a critique of capital. In his theorization, the material conditions, the capitalist mode of production based on the sale of wage labor as a commodity and its resale as exchange value, constituted the person as a thing, thwarting the move from immanence to transcendence and the possibility of human fulfillment.

Marx's 1844 *Manuscripts* showed that the worker no longer controlled the labor process; she owned neither her materials nor her tools, or the products that she produced for the market. While Hegel's Slave could acquire self-consciousness through transforming nature, even though the process and the products were appropriated by the Master, under the capitalist system, this was impossible. Arguing that Master and Slave were to be understood as classes—those who owned capital, the bourgeoisie, and those who sold their labor power, respectively—Marx shows that workers were rendered powerless by the very economic system they made possible. With the externalization of the products of their own labor, the capitalist system then stood outside the workers, refluxed back upon them as an alien force, turning them into powerless objects. With their labor reduced to costs of production, wealth was produced for the capitalist and immiseration for the objectified workers. Consequently, they were estranged from their work and its products, their selfhood and its potential, isolated from each other and bereft of their very humanity (species-being, *Gattungswesen*) as elaborators of nature.

In sum, human growth, freedom, creativity, and community were frustrated by "commodity fetishism" integral to capitalism. The social relationship of capitalist and worker was reified, congealed within commodities, and sold for their exchange value. Products/commodities were then endowed with surplus value, while the concrete social relationship between those who own capital and those whose work produces the appropriated surplus value is intentionally hidden.

György Lukács

For Lukács (1885–1971), the eminent Hungarian Marxist philosopher, reification (from the Latin, *res,* "thing") was a special form of alienation for understanding bourgeois society. While the *1844 Manuscripts* had not yet been found when he wrote *History and Class Consciousness,* his close reading of Marx on "commodity fetishism" revealed the relation of alienation to commodification and how, in turn, reification shaped both the structure of capitalist society and its subjective moments. Describing commodity structure as "a relation between people," he insisted that it

> takes on the character of a thing and thus acquires a "phantom objectivity," an autonomy that seems so strictly rational and all-embracing as to conceal every trace of its fundamental nature: the relation between people. . . . The reification produced by commodity relations assumes decisive importance both for the objective evolution of society and for the stance

> adopted by men towards it . . . the commodity become[s] crucial for the subjugation of men's consciousness to the forms in which this reification finds expression and for their attempts to comprehend the process or to rebel against its disastrous effects and liberate [themselves] from servitude to the "second nature" so created. (pp. 83–86)

Reification, in which social relations become objectified, is embedded within bourgeois ideology and consciousness, thwarting the ability of the working class to articulate its own standpoint—notwithstanding how they produced the very wealth that was the foundation of capitalist society and the basis of their reification. Nor could they become aware of themselves as the unified subject–object of history.

Frankfurt School

In the work of the Frankfurt School, alienation, qua the reduction of workers to reified objects with attenuated social bonds, meant that people experienced themselves as socially fragmented, powerless objects whose externalized work activities and their social consequences have become the dominant forces in their lives. They were bereft of experiencing themselves as creative agents shaping their own destinies, they experienced the world as "a complicated social machine to administer the technical machine [Man has] built. The more powerful and gigantic the forces are which he unleashes, the more powerless he feels himself as a human being" (Fromm, 1955, p. 124). To resolve these contradictions of capitalism, such people were disposed to seek comfort in fascism/totalitarianism (this is what Theodor Adorno called the authoritarian personality).

Instrumental reason, the dominant ideology of capitalism, aided and abetted by a mass-mediated popular culture of deception and distraction, has served as a technique to produce and sustain alienation through the domination of consciousness. Moreover, the post–World War II expansion of capitalism colonized desire and consciousness in a society where more and more people based their very identities and lifestyles on consumption. For theorists like Herbert Marcuse, alienation was no longer based on the production of commodities but was seen in "one-dimensional thought," consumerism, and "repressive desublimation."

Sociology

Empirical Social Research

While philosophers engaged in various forms of critique, sociologists, informed by Max Weber's theories of demystification and entrapment by reason and by Émile Durkheim's discussions of anomie, undertook empirical investigations of alienation. Robert Merton's classical analysis of anomie as the degree of concordance of means and ends encouraged a number of empirical studies. Following on M. Seeman's literature review in which he found that alienation had been treated in sociological theory as meaninglessness, powerlessness, normlessness, social isolation, and self-estrangement, several researchers operationalized these dimensions of alienation in scales of various degrees of sophistication. Robert Blauner conceptualized alienation by examining four kinds of work settings—textiles, automobile assembly, lithography, and chemical process—to explore alienation qua creativity and agency in the work process, and workers' cohesion and solidarity. He found that alienation was inversely correlated with workers' skillfulness and cohesion. Recent studies have researched alienation among immigrants, among adolescents, and in families, as well as in education. Critics of this empiricist trend argue that such research reduces capitalism to a social psychology of individual differences and its positivistic empiricism is itself an expression of alienation.

Postmodernism and the Return of the Alienated Actor

In the late 1970s and 1980s, critics such as the French philosopher Jean-François Lyotard and the French sociologist, philosopher, and cultural critic Jean Baudrillard argued that having reached the postmodern age, the grand narratives of modernity, such as those of Marx or Weber, were no longer relevant. The massive expansion of consumerism and domination by the mass media, along with the demise of socialist regimes, signaled the downfall of the political and the advent of a world of decentered subjects exposed to endless simulations. Concern with alienation waned as foci shifted to cultural studies and the humanities, privileging local knowledge. For the American Marxist literary and social theorist Fredric Jameson, the ascent of postmodernism with its endless media spectacles is the cultural

expression of a ruthless globalized capitalism. While hyperconsumption provided both profits for global capital and its legitimating ideology, it colonized consciousness, transformed desire, and resulted in the proliferation of "shopping mall selfhood."

In early capitalism, Marx saw alienation as a condition for the production of commodities; in late capitalism, commodified selfhood is objectified, powerless, estranged from its own potential. The only genuine bases of community are shared consumer tastes, membership in mass audiences, and affiliation with fandoms—all of which are expressions of alienation.

Meanwhile, with wages stagnant, extravagant consumer tastes were financed largely through credit. By the late 1980s and 1990s, the implications of globalization were becoming more evident as manufacturing jobs were increasingly either automated or moved to the developing countries. Indeed, consumer debt, intertwined with consumption-based lifestyles, became another source of alienation. Although this was evident to a number of economists, sociologists, and political scientists, it would take the great meltdown of 2007 for debt-based consumerism to be seen as the catalyst for widespread alienation. Some genres of popular culture embraced transgressive themes and grotesque aesthetics—"carnivalization"—in which various moral transgressions can be seen as resistance via a critique of capitalist alienation and dehumanization of rationality. A significant sign of the trend is the spread of nihilism in such forms of popular music as punk and heavy metal and their offshoots. Such subcultures would seek to resist, if not overcome, alienation through the embrace of an alternative cultural agenda, but ultimately, given the role of the culture industries, the seeming resistance sustains the reproduction of capital.

Globalization/Economic Stagnation/Crisis

With the implosion of the European and American economies, as millions of people lost their jobs and often their homes, alienation became palpable in anger and *ressentiment* (to use Nietzsche's term). Capitalism could no longer promise upward mobility, affluence, and consumerism to cloak its underlying effects. A typical reaction was White right-wing mobilization in the United States, with the Tea Party as a voice of discontent not unlike the European fascist mobilizations of the 1920s and 1930s. These movements generally tend to express anger against existing elites as well as xenophobia toward those seen as "outsiders."

But some mobilizations are progressive, among them the World Social Forum, in which activists, embracing a variety of causes, gather from all over the world to forge networks and coalitions that would make a world of equality and humanity possible. In the spring of 2011, after many years of oppressive political leadership, economic stagnation, and shortages of basic commodities, vast numbers of people in the Middle East mobilized to rise up against oppressive governments, in attempts to overcome powerlessness and alienation.

Conclusion

When Hegel criticized Immanuel Kant's notion of Reason as static and ahistorical, he introduced the concept of alienation to explain the movement of spirit (*Geist*) through history. Moreover, the center of his phenomenology was the master's appropriation of the labor and recognition of the bondsman. But when Marx appropriated the concept of alienation, he offered a critique of capitalist domination in societies based on wage labor congealed within commodities produced for exchange. He claimed that the contradictions of capitalism would lead to a society based not on necessity but on freedom, community, and the full realization of species-being. Little did Marx anticipate that his insights would not only foster revolutions but also generate a great deal of sociological theory and research. His insights on alienation continue to help us understand the consequences of global capitalism in its consumerist moment, with its popular culture and its many reactionary and progressive mobilizations.

Lauren Langman and
Devorah Kalekin-Fishman

See also Capitalism; Dialectic, in the Social Sciences; Feminism: Schools of Thought; Frankfurt School and Critical Social Theory; Idealism; Ideology; Marxism and Social/Historical Explanation; Nihilism; Oppression; Reification

Further Readings

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality.* New York, NY: Harper & Row.

Anderson, G. (2005, May). *Indices and tests for alienation based upon gini type and distributional overlap measures*. Paper presented at the Economics Department, University of Toronto, Ontario, Canada.
Blauner, R. (1964). *Alienation and freedom: The factory worker and his industry*. Chicago, IL: University of Chicago Press.
Dean, D. G. (1961). Alienation: Its meaning and measurement. *American Sociological Review, 26*, 753–758.
Feuerbach, L. (1989). *The essence of Christianity* (G. Eliot, Trans.). Amherst, NY: Prometheus Books. (Original work published 1854)
Finifter, A. (1970). Dimensions of political alienation. *American Political Science Review, 64*(2), 389–410.
Fromm, E. (1955). *The sane society*. New York, NY: Holt Rinehart & Winston.
Hegel, G. W. F. (1977). *The phenomenology of spirit* (A. V. Miller, Trans.). Oxford, England: Oxford University Press. (Original work published 1807)
Israel, J. (1971). *Alienation: From Marx to modern sociology: A macrosociological analysis*. Boston, MA: Allyn & Bacon.
Jameson, F. (1984). *Postmodernism, or the cultural logic of late capitalism*. Durham, NC: Duke University Press.
Jimenez, T. G. (2010). They hatch: The alienation of the colonial American subject in Toni Morrison's *A Mercy*. *Berkeley Undergraduate Journal, 22*(2). Retrieved from http://escholarship.org/uc/item/0ds156cw
Kelly, J. B., & Johnston, J. R. (2005). The alienated child: A reformulation of parental alienation syndrome. *Family Court Review, 39*(3), 249–266. doi:10.1111/j.174–1617.2001.tb00609x
Langman, L., & Kalekin-Fishman, D. (2005). *The evolution of alienation: Trauma, promise and the new millennium*. Lanham, MD: Rowman & Littlefield.
Lee, A. M. (1972). An obituary for "alienation." *Social Problems, 20*(1), 121–127.
Lukács, G. (1971). *History and class consciousness* (R. Livingstone, Trans.). Cambridge: MIT Press. (Original work published 1923)
Lyotard, J.-F. (1984). *The postmodern condition: A report on knowledge*. Minneapolis: University of Minnesota Press. (Original work published 1970)
Marx, K. (1988). *Economic and philosophic manuscripts of 1844* (M. Milligan, Trans.). Amherst, NY: Prometheus Books. (Original work published 1844)
Merton, R. K. (1938). Social structure and anomie. *American Sociological Review, 3*(5), 672–682.
Mészàros, I. (1971). *Marx's theory of alienation*. London, England: Merlin Press.
Ollman, B. (1985). *Alienation: Marx's conception of man in capitalist society*. Cambridge, England: Cambridge University Press.
Overend, T. (1975). Alienation: A conceptual analysis. *Philosophy and Phenomenological Research, 35*(3), 301–322.
Sayers, S. (2011). *Marx and alienation: Essays on Hegelian themes*. London, England: Palgrave Macmillan.
Schacht, R. (1994). *The future of alienation*. Springfield: University of Illinois Press.
Seeman, M. (1959). On the meaning of alienation. *American Sociological Review, 24*, 783–791.
Seeman, M. (1971). Alienation: A map. *Psychology Today, 10*, 84–87.
Slater, M. D. (2003). Alienation, aggression, and sensation-seeking as predictors of adolescent use of violent film, computer, and website content. *Journal of Communication, 53*(1), 105–121.

Allais Paradox

The Allais paradox, discovered by Maurice Allais, provides an example in Decision Theory of preferences that violate the most widely accepted normative theory of decision making, expected utility theory. This entry briefly explains expected utility theory, along with the paradox, and describes responses to the paradox.

The Paradox

Consider a decision maker choosing between lotteries, that is, probability distributions over outcomes. According to expected utility theory, as long as the decision maker is rational, her preferences can be represented by a utility function of outcomes with the property that of any two lotteries she prefers the lottery with the higher expected utility value. The idea that decision makers maximize expected utility—the "expected utility hypothesis"—was put forth in part to account for the fact that many decision makers are risk averse in the sense that they would rather have, for example, a sure-thing amount of money than a lottery with the same average monetary value. While such behavior is not consistent with maximizing expected *monetary* value, it is consistent with maximizing expected utility, relative to a concave utility function of money.

The link between a decision maker's preferences and her utility is cemented by "representation theorems": These theorems show that being (representable as) an expected utility maximizer is equivalent to having preferences that satisfy particular axioms. One of the earliest and most influential axiomatizations is that of John von Neumann and Oskar Morgenstern. The axioms of von Neumann and Morgenstern's theorem—and those employed in representation theorems in general—seem to many to be requirements of rational preferences.

The Allais paradox is a counterexample to the expected utility hypothesis. Allais asks us to consider the following choice scenarios. First, we are asked whether we prefer Situation A or Situation B.

Situation A: \$100 million for certain.

Situation B: a 10% chance of \$500 million, an 89% chance of \$100 million, a 1% chance of nothing.

We are then asked whether we prefer Situation C or Situation D.

Situation C: an 11% chance of \$100 million, an 89% chance of nothing.

Situation D: a 10% chance of \$500 million, a 90% chance of nothing.

Allais hypothesized that most people strictly prefer A to B and also strictly prefer C to D, on the grounds that in the first choice scenario the advantages of certain gain from A outweigh the perhaps higher but uncertain gain from B but that in the choice between C and D the much higher gain outweighs the slightly higher probability of a much lower gain. This pattern of preferences has been confirmed experimentally.

As mentioned, this pattern of preferences violates the expected utility hypothesis: There is no possible assignment of utility values to \$0, \$100 million, and \$500 million such that A has a higher expected utility than B and D has a higher expected utility than C. (One can strictly prefer A to B and C to D, or one can strictly prefer B to A and D to C, but no other combination of strict preference satisfies the expected utility hypothesis.) The particular axiom that these preferences violate is von Neumann and Morgenstern's independence axiom.

Responses

Broadly speaking, there are two ways to take a Decision Theory: as an analysis of the canons of instrumental rationality ("normative" Decision Theory) or as a description of actual people's preferences ("descriptive" Decision Theory). For normative decision theorists, the standard choices in Allais's example are "paradoxical" in that they seem rational to many people and yet they violate the dictates of the expected utility theory, which seems to correctly spell out the requirements of rationality. For descriptive decision theorists, the Allais choices aren't so much paradoxical as they are a counterexample to the idea that expected utility theory is the correct descriptive theory.

Descriptive theorists have responded to the paradox by formulating alternative theories that are compatible with the Allais choices.

For normative decision theorists, there are three ways to respond to the paradox. The first is to claim that contrary to initial appearances, the Allais choices are simply irrational: Although many people unreflectively have the standard Allais preferences, once an individual sees that her preferences violate the independence axiom, she ought to reevaluate her preferences and bring them in line with expected utility theory.

A second response to the paradox is to claim that contrary to initial appearances, the Allais choices do satisfy the expected utility hypothesis and the apparent conflict is due to the fact that the choices have been underdescribed in the initial setup. Since decision makers prefer A to B on the grounds that A yields \$100 million *for certain* or on the grounds that an individual who takes B and ends up with nothing will feel *regret,* this response claims that the actual outcomes in the problem are not simply monetary amounts but also include the decision maker's feelings about getting those outcomes in each situation. Since it is possible to assign utility values to, for example, the outcomes \$0, \$100, \$500, and \$0 *with regret* such that the Allais choices maximize expected utility, they do not violate expected utility after all.

Finally, normative decision theorists might respond to the paradox by denying the expected utility hypothesis and arguing that expected utility theory is inadequate as a theory of rationality. This response claims that the Allais choices genuinely

violate the theory and that they are nonetheless rational. Theorists advocating this response may draw on the aforementioned descriptive theories and argue that the preferences of the decision makers they describe are in fact rational.

Lara Buchak

See also Agnotology, Ignorance, and Uncertainty; Decision Theory; Folk Psychology; Formal Epistemology; Preference; Rational Expectations; Risk

Further Readings

Allais, M. (1979). The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American school. In M. Allais & O. Hagen (Eds.), *Expected utility and the Allais paradox* (pp. 27–145). Dordrecht, Netherlands: D. Reidel. (Original work published 1952)

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–292.

Savage, L. (1954). *The foundations of statistics*. New York, NY: Wiley.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Analytical Marxism

Analytical Marxism includes philosophers and social scientists who use the methods of analytical philosophy, modern economic theory, and sociology to clarify or reconstruct ideas central to Marxist social theory while discarding other, less well-grounded aspects of it. This entry examines the background and assesses the success of this school within Western Marxism. It also shows important epistemological lessons social science can draw from the methodological debates on Marxist social explanation that took place within this school.

Early Cold War critiques of Marxism claimed that Marx's philosophy of history and society was both conceptually and theoretically confused. Analytical Marxism responded by promising to reconstruct Marx's theories so that they were clear, rigorous, and scientifically supported, or discard them if this failed. Analytical Marxism also was a response to interpretations of Marx's social theory propagated by Marxist political movements and those influenced by the German idealist tradition within Western continental philosophy. In this respect, analytical Marxism is primarily a phenomenon of the philosophy of Marx in the English-speaking world, self-consciously exhibiting the "common sense" of English philosophy.

Analytical Marxism progressively discarded or reconstructed some key concepts and theoretical premises of Marxism, thereby raising the question of whether Marx's social theory could bear rigorous scrutiny. It more or less sympathetically attempted to reconstruct Marx's theory of classes, theory of historical materialism, theory of exploitation, and theory of capitalism. There were some dismissive overall surveys of Marx's social theory, such as Jon Elster's *Making Sense of Marx* (1985).

Analytical Marxists largely reject two aspects of Marx's theory that he considered among his most significant contributions. One was his use of a critique of classical political economy to reveal the conflict between classes at the foundation of capitalist society. The other was his claim that this conflict would result in a revolutionary transformation of capitalism into socialism. With regard to the first, many analytical Marxists dismiss Marx's analysis of class and exploitation because it relies on an outdated labor theory of value and a distinction between labor and labor power. Rejection of the labor theory of value led most to dismiss Marx's measure of the rate of exploitation as surplus value divided by the value of labor power, though this measure was sometimes equated with Marxian exploitation itself. Reliance on neoclassical economic theory also led to rejection of Marx's picture of capitalist exploitation as the basis of capitalist industry, which Marx claimed was concealed by superficial fairness in the way free laborers and capitalists agree on the terms and conditions of employment.

Analytical Marxists are divided among themselves between those who think that socialism is possible and those who think socialism can only be a merely utopian ideal of working-class politics, which realistically should only aspire to reforms within the ambit of welfare capitalism. Elster claims that Marxist political movements face a dilemma. In less developed capitalist societies, revolutionary mass action is possible, but what Marx claims to

be necessary conditions for socialism are missing. In developed capitalist societies, objective social conditions for socialism might be present, but the working class does not want change strongly enough to form revolutionary working-class political movements.

Some analytical Marxists use mathematical models of rational, self-interested choice by individuals to question whether revolutionary struggles can be sustained if participants are not subject to coercive discipline, like that imposed by what the Russian revolutionary Vladimir Ilyich Lenin called a "vanguard" party. Their analysis also raises the more general question of whether free collective action is possible and whether socialism can be seen as a free association of workers, as Marx saw it. The alternative takes socialism as essentially a planned economy, which has been shown in practice to be very inefficient and corrupt, and so hardly worthy of support.

One characterization of analytical Marxism is that it rejects any sharp separation of Marxist methodology of science from methods used in the social sciences more generally. In itself, this should be acceptable to other interpreters of Marx. A problem arises when some leading analytical Marxists reject some of Marx's ideas with as little care as some other Marxists show when they slavishly accept them. In these cases, little attempt is made to clarify what Marx actually thought or, where necessary, reconstruct his ideas. Dialectics and dialectical method are dismissed as "bullshit" rather than clarified to see whether they have substance or are worth reconstructing.

Although John Roemer provides a striking and original account of class and exploitation, he constructs more or less parallel ideas to those of Marx rather than providing a careful interpretation of Marx's social theory. Roemer shows that you can profit from lending money to workers. He also shows that owners of better-quality land can exploit owners of poorer land. Roemer thus thinks Marx is misguided in claiming that capitalist control over the duration, intensity, and productivity of the labor of their employees is crucial to capitalist exploitation. However, Roemer is too intent on leaving Marx behind to notice that Marx may accept his claims about exploitation through moneylending and control of more productive land. Marx has a general notion of exploitation, where some live off the forced labor of others, and a specific version of this within a capitalist economy. Marx thus may allow forms of exploitation other than a specifically capitalist form. Marx never simply equates exploitation with a transfer of surplus value or surplus labor, as Roemer claims.

Analytical Marxism faded away after the fall of the Soviet Union, although Gerald Cohen continued until his death in 2009 to argue strongly for the justice of socialism, while tackling John Rawls's theory of justice.

Ian Hunt

See also Dialectic, in the Social Sciences; Game-Theoretic Modeling; Marxism and Social/Historical Explanation; Marxist Economics; Mathematical Models, Use in the Social Sciences; Rational Choice and Political Science

Further Readings

Cohen, G. A. (2000). *Karl Marx's theory of history: A defence* (Expanded ed.). Oxford, England: Clarendon Press.

Elster, J. (1985). *Making sense of Marx*. Cambridge, England: Cambridge University Press.

Hunt, I. (1998). A dialectical interpretation and resurrection of historical materialism. In R. Panasiuk & L. Nowak (Eds.), *Marx's theories today* (pp. 153–193). Amsterdam, Netherlands: Rodopi.

Larrain, J. (1986). *A reconstruction of historical materialism*. London, England: Allen & Unwin.

Roberts, M. (1996). *Analytical Marxism: A critique*. New York, NY: Verso.

Roemer, J. E. (1982). *A general theory of exploitation and class*. Cambridge, MA: Harvard University Press.

Roemer, J. E. (Ed.). (1986). *Analytical Marxism*. Cambridge, England: Cambridge University Press.

Roemer, J. E. (1988). *Free to lose: An introduction to Marxist economic philosophy*. Cambridge, MA: Harvard University Press.

Steedman, I. (1977). *Marx after Sraffa*. London, England: New Left Books.

Wright, E. O. (1997). *Classes*. New York, NY: Verso. (Original work published 1985)

Wright, E. O., Levine, A., & Sober, E. (1992). *Reconstructing Marxism: Essays on explanation and the theory of history*. New York, NY: Verso.

Analytical Sociology and Social Mechanisms

Analytical sociology is a reform movement within sociology and social theory. Its identity is not based on a common object of study, a shared historical

tradition in sociological theory, or an empirical research method. Rather, it is founded on the idea that the social sciences should do more than describe and classify social processes. According to analytical sociologists, the primary epistemic aim of the social sciences should be the *causal* explanation of social phenomena. Sociological theory should aim to develop clear and precise accounts of the *social mechanisms* by which the intentional activities of social agents bring about social phenomena.

The historical roots of analytical sociology can be traced back to the works of late-19th-century and early-20th-century pioneering social theorists such as Alexis de Tocqueville and, later, Max Weber and to prominent mid-20th-century sociologists such as the early Talcott Parsons and Robert K. Merton. Among contemporary social scientists, Jon Elster, Raymond Boudon, Thomas Schelling, and James Coleman have profoundly influenced the analytical approach. Although they are rather different types of scholars, they complement each other in important ways, and they all share a commitment to precise, abstract, and action-based explanations.

The current generation of analytical sociologists is building upon the foundations laid by these authors, aiming to develop an analytical middle-range approach to sociological theory that avoids the somewhat empiricist and eclectic tendencies of Merton's original middle-range approach. Examples of recent work in this tradition can be found in anthologies edited by Hedström and Swedberg (1998), by Hedström and Bearman (2009), and in a monograph by Hedström (2005).

Social Mechanisms

The core idea of analytical sociology is that sociological theory explains by specifying the *causal mechanisms* by which social phenomena are brought about. The specifically social causal mechanisms are called *social mechanisms*.

Analytical sociologists are dissatisfied with the traditional covering-law account of explanation, which has a large number of philosophical problems and embarrassing counterexamples. It has also provided justification for the use of "black box" explanations in the social sciences, as it does not require that the mechanism linking *explanans* and *explanandum* be specified for an acceptable explanation to be at hand. This omission has given leeway for sloppy scholarship. In contrast to this fundamentally empiricist account of explanation, analytical sociologists require that explanations should articulate causal mechanisms rather than simply subsume phenomena under empirical generalizations.

Apart from providing explanatory understanding as to why the dependency holds, the information about the causal mechanisms also provides justification for causal claims. Distinguishing between real causal claims and spurious statistical associations is a major challenge in the social sciences. The real causal dependences are transmitted via causal processes, and the search for causal mechanisms directs attention to these processes.

The idea of causal mechanisms is related to broader ideas about the growth and organization of scientific knowledge. In mechanism-based accounts, scientific knowledge is embedded in mechanism schemes and not in empirical generalizations, as in more traditional empiricist accounts. According to this view, scientific knowledge expands by adding items to, or improving upon, items already present in the toolbox of possible causal mechanisms. Understanding accumulates as the knowledge of mechanisms gets more detailed and the number of known mechanisms increases.

The mechanism idea is important in a highly specialized and fragmented discipline like sociology. Although empirical data, research methods, and substantial theories differ from one subfield of sociology to another, the general ideas about possible causal mechanisms are something these fields could share and thereby benefit from each other's work. In this vision, sociological theory provides a set of explanatory tools that can be employed and adapted to particular situations and explanatory tasks. The mechanisms are (semi)general in the sense that most of them are not limited to any particular application. For example, the same type of mechanism can be used for (partially) explaining residential segregation and success in cultural markets.

This mechanism-based vision of knowledge has much in common with Robert K. Merton's idea of sociological theories of the middle range. A theory of middle range is a clear, precise, and simple type of theory, which can be used for partially explaining a range of different phenomena but which makes no pretense of being able to explain all social phenomena and is not founded upon any form of extreme reductionism in terms of its *explanans*. Middle-range theories isolate a few explanatory factors that explain important but delimited aspects of the outcomes to be explained.

Social *Explananda* and Individual Action

The main focus of analytical sociology is on social (as distinct from psychological) *explananda* that are not definable by reference to any single member of the collectivity. Important examples of such properties include the following:

- Typical actions, beliefs, or desires among the members of a collectivity
- Distributions and aggregate patterns, such as spatial distributions and inequalities
- Topologies of networks that describe relationships between the members of a collectivity
- Informal rules or social norms that constrain the actions of the members of a collectivity

A key insight of analytical sociology is that explanations that directly relate macrovariables to each other are unsatisfactory. They do not specify the causal mechanisms by which the changes in one macro variable bring about changes in another. Deeper explanatory understanding requires opening up the black box and finding the causal mechanisms that have generated the macro-level pattern. Rather than analyzing relationships between phenomena exclusively on the macro level, one should identify the *situational mechanisms* by which social structures constrain individuals' action and cultural environments shape their desires and beliefs, describe the *action formation mechanisms* according to which individuals choose how to act, and specify the *transformational mechanisms* by which individuals, through their actions and interactions, generate various intended and unintended social outcomes. Only by understanding the whole chain of situational, action formation, and transformational mechanisms have we made sense of the observed macro-level relationship.

This argument does not imply commitment to the doctrine of methodological individualism. Most formulations of methodological individualism are much stronger than is required by the mechanism-based perspective. Some form of *structural individualism* is sufficient for the purposes of mechanism-based explanations of social phenomena. Structural individualism is a doctrine according to which all social facts, their structure and change, are in principle explicable in terms of individuals and their properties, actions, and relations to one another. Structural individualism differs from most formulations of methodological individualism by emphasizing the explanatory importance of relations and relational structures. It does not require that all explanatory facts are facts about individual agents in the strict sense: Facts about topologies of social networks; about distributions of beliefs, resources, or opportunities; and about institutional or informal rules and norms can have a significant role in mechanism-based explanations.

Historically, analytical sociology has some of its roots in the tradition of rational choice sociology. However, current analytical sociology does not have any special commitment to the assumptions of rational choice theory. For this reason, it is a mistake to take analytical sociology as rebranded rational choice sociology. Rather than relying on some preconceived ideas about human motivation or cognitive processing, the mechanism-based perspective suggests that our account of human agency should be based on the findings and theories of psychological and cognitive sciences. The explanatory agenda of mechanism-based social science has some methodological implications, however. As understanding of complex phenomena is only possible in a piecemeal way, we must abstract away from many details of human mental life. Only those aspects of cognition that are relevant for the explanatory task at hand should be included in the explanation; the explanatory task thus determines how rich the psychological assumptions must be.

So, although the mechanism-based approach emphasizes the importance of action in the explanation of social phenomena, it does not subscribe to an axiomatic vision according to which a specific action theory should be used for all purposes. As different theories of action emphasize different aspects of human action, the choice between them should be made on methodological grounds. In some modeling contexts, rational choice theory is a natural choice, while in contexts where the emphasis is on the habitual dimensions of human actions, the analytical sociologist might choose to employ some psychological dual-process theory. For many social-scientific purposes, a relatively simple desire–belief–opportunity model will be sufficient. This simple theory provides a building block for accounts of social mechanisms of interaction through which the actions of some actors may come to influence the beliefs, desires, opportunities, and actions of others.

Agent-Based Modeling

A recent development in analytical sociology has been the use of agent-based computer simulations. Much of the development of mechanism-based knowledge consists of developing "how-possibly" explanation schemes. These schemes are not intended to explain any particular empirical facts directly but to provide a general understanding of how things could work. Given the limitations of experimental methods and the complexity of social phenomena, computer simulations are important for this kind of endeavor.

The promise of agent-based simulation is based on the fact that the dynamics observed at the social level typically are complex and hard to understand, but often it is possible to describe the basic cogs and wheels of these social processes with rather simple models. Macro-level outcomes and relationships tell us very little about why we observe the macro-level outcomes and relationships we do observe. Only by explicitly considering the micro-level actions and relations and how they unfold over time can macro-level outcomes be explained. This basic insight is at the heart of analytical sociology: To understand collective dynamics, we must study the collectivity as a whole, but we must not study it as a collective entity.

One important feature of agent-based simulations is that they do not impose any a priori constraints on the mechanisms assumed to be operating. Unlike rational choice theory, agent-based modeling is not based on any specific theory of action or interaction. It is a methodology for deriving the social outcomes that groups of interacting actors are likely to generate, whatever the action logics or interaction structures may be. Agent-based simulations should not only be regarded as a tool for theoretical exploration, however. Empirically calibrated agent-based models make it possible to integrate theoretical ideas with the results of empirical research.

Conclusion

Mechanism talk is becoming increasingly popular in the social sciences. Many sociologists use the word *mechanism* in a loose sense without any commitment to the type of mechanism-based explanatory strategy focused upon here. Underlying the mechanism-based approach is a commitment to *realism:* The explanations should reflect the causal processes responsible for the phenomenon to be explained. This requires stringency in theoretical practice and imagination in the design of empirical research. It also implies that social scientists should pay attention to what mechanisms are made of rather than treating them as black boxes.

Peter Hedström and Petri Ylikoski

See also Agent-Based Modeling and Simulation in the Social Sciences; Causation in the Social Sciences; Economic Sociology; Explanation, Theories of; Individualism, Methodological; Mathematical Models, Use in the Social Sciences; Mechanism and Mechanismic Explanation; Models in Social Science

Further Readings

Coleman, J. S. (1990). *Foundations of social theory.* Cambridge, MA: Belknap Press.

Hedström, P. (2005). *Dissecting the social: On the principles of analytical sociology.* Cambridge, England: Cambridge University Press.

Hedström, P., & Bearman, P. (Eds.). (2009). *The Oxford handbook of analytical sociology.* Oxford, England: Oxford University Press.

Hedström, P., & Swedberg, R. (Eds.). (1998). *Social mechanisms: An analytical approach to social theory.* Cambridge, England: Cambridge University Press.

Hedström, P., & Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Sociology, 36,* 49–67.

Macy, M., & Flache, A. (2009). Population dynamics from the bottom up: Agent-based models of social interaction. In P. Hedström & P. Bearman (Eds.), *The Oxford handbook of analytical sociology* (pp. 245–268). Oxford, England: Oxford University Press.

Merton, R. K. (1967). *On theoretical sociology.* New York, NY: Free Press.

Udéhn, L. (2001). *Methodological individualism: Background, history and meaning.* London, England: Routledge.

Analytic/Synthetic Distinction

The analytic/synthetic distinction was motivated historically by a number of issues. Immanuel Kant explicitly drew the distinction as it has been best known since. Synthetic statements are often understood as factual statements known by experience. In contrast, some statements seem knowable by reason

alone, without appeal to sense experience. Such statements are called a priori. Some statements seem necessary, not merely contingently true. If the necessity of a statement cannot be known by appeal to experience, then the fact that *p* is necessary seems to be knowable, if at all, a priori. Various statements seem to be uninformative or trivial in a peculiar way. Some examples are obvious truths of logic, such as "If p, then p," or well-known truths concerning abbreviational definitions, such as "Vixens are female foxes." David Hume distinguished "relations of ideas" from "matters of fact" in a way motivated by this third distinction, although it is not altogether clear why "relations of ideas" should be knowable **a priori**, or trivial/uninformative, or necessary. It is clear that epistemic distinctiveness (experiential transcendence) played a major role historically in motivating the distinction.

The Logical Empiricist View

The analytic/synthetic distinction was a central pillar of the form of empiricism that came to dominate the first half of the 20th century, logical empiricism. According to logical empiricists, any statement whose truth is knowable, but not knowable on the basis of sense experience, is analytic. Any statement neither analytic nor synthetic is "metaphysical" in a pejorative sense. Logical truths, truths of mathematics, and various seeming necessary truths, such as color exclusion claims (that nothing can be two different colors at the same time), were particularly difficult cases for the logical empiricists, who wanted to assimilate all nontrivial or informative statements to the "synthetic." Their picture of scientific knowledge required that there be no "substantive" a priori knowable truths that are not analytic.

Alfred J. Ayer's famous book, *Language, Truth, and Logic,* presented a description of our epistemic circumstances as understood by logical empiricists. Ayer assimilated the necessary, the a priori, and the "trivial" or "uninformative" to the same category of statements, the analytic ones. Since a modal feature such as necessity seems to transcend what is determinable empirically, as Hume and other empiricists had argued, and we take some statements of science to be necessary, knowledge of such statements and/or their modal status were thought to require appeal to analyticity, and similarly for other apparent necessities such as color exclusion. Both mathematics and logic seem to be knowable, if at all, without appeal to sense experience, at least in the sense that our justification for believing such truths does not seem to appeal to sense experience. Ayer argued that no empirical data would be taken to undermine our confidence in, say, $2 + 3 = 5$, and that no empirical data could be taken to confirm these statements either. Rather than arguing that our arithmetical beliefs are unjustified, logical empiricists preferred to assimilate these statements to the class of analytic statements.

Rudolf Carnap emerged as the most prominent of the logical empiricists, and he developed various accounts of the workings of language and of scientific methodology that crucially involved the notion of an analytic statement. It is Carnap's developed views, as well as the logical empiricist views expressed by Ayer, that generated some of Quine's most powerful objections to the distinction.

Quine's "Two Dogmas of Empiricism"

Willard Van Orman Quine wrote a number of papers attacking the analytic/synthetic distinction, the best known of which is "Two Dogmas of Empiricism (TD)." In that paper, he raised a number of objections to analyticity. One objection is a form of circularity objection. Quine noted that various attempts to clarify the notion of analyticity appealed to other notions such as necessity, semantic rule, or synonymy. According to Quine, all of these notions, to the extent that they seem interdefinable with "analytic," are equally problematic. This class of interdefinable notions includes meaning and synonymy, both of which he rejects as obscure or worse.

A second part of Quine's TD argues that no statement within science is "immune to revision" on the basis of experience. According to Quine, even truths of logic seem alterable on the basis of empirical facts (e.g., "quantum logic," which rejects the distributive law of classical logic in response to experimental data pertaining to the quantum realm). Mathematical statements are empirically revisable as well (as observed, they claim, in the transition from Euclidean to non-Euclidean geometries). This argument is given in order to undermine another potential account of analytic statements, that they are those statements that are taken to be "confirmed" relative to any experience.

Around two decades later, Gilbert Harman summarized the state of the arguments concerning analyticity, extending and improving Quine's original

arguments to better address some of the objections raised against it by Paul Grice, Peter Strawson, Carnap, and others. Harman argued that "analytic" was akin to "witch," in the sense that we now think that there are no witches because appeal to witchhood does no helpful explanatory work and that we should similarly reject analyticity because it does no helpful explanatory work either.

Some Responses to TD

Some early responses to TD raised difficulties for Quine's position. For example, Quine claims in TD that stipulative definition for the purposes of abbreviation yields unproblematic analytic statements. Grice and Strawson argue that Quine cannot consistently accept such stipulations as analytic. If there are some clear cases of analytic statements, then the notion must not be incoherent or unintelligible, contrary to the central conclusion of Quine's TD.

Many commentators have argued that Quine's circularity arguments are unconvincing. Other fundamental notions, such as one of Quine's favorites, truth, seem indefinable. This should not cast doubt on a notion's coherence or as to whether the concept has instances. Furthermore, as Glock points out, Quine's argument would appear to stack the odds unfairly against the possibility of explication. Quine treats any concept used to elucidate analyticity to be equally suspicious, to suffer "guilt by association." It is difficult to see how any elucidation of any concept could ever be given under the constraint that any proposed clarification is taken to cast doubt on the concepts employed in the attempted elucidation.

One of the central points that Grice and Strawson make is that philosophers seem capable of classifying an indefinite and open-ended number of statements as either analytic or not, with considerable agreement. This shows, they think, that there must be some intelligible notion that is implicitly shared by philosophers, even if they cannot yet give an explicit and clear elucidation of the notion.

In response to this second objection, Harman argues that the fact that people agree by and large on their classifications does not show that there are, in fact, instances of a concept. For example, it may have been the case that in Massachusetts in the 1600s people would have substantially agreed on how to classify women as witches or nonwitches. That fact does not show that there were actually witches.

Another sort of objection that has been made is that if Quine is correct in his arguments against meaning and synonymy, there is no distinction between change in meaning and change in theory. But intuitively, there seem clear cases in which a word has changed its meaning, as opposed to merely changing our views concerning the term. A Quinean response to this objection grants that there seem to be such different cases but attributes the difference to our inclination to use two different sorts of "translation schemes" in the two cases. Sometimes we translate the same word differently across time into our own language, whereas other times we translate the word, as used across time or community, into the same word of our current language. The notion of a good translation scheme is not, according to Quineans, to be understood as requiring preservation of meanings, since there are no meanings and there is no such thing as synonymy. Instead, good translation schemes are the ones that have some pragmatic virtues, including preserving of belief (treated behavioristically as "dispositions to assent to sentences") as well as preserving "obviousness."

As to Harman's concern that analyticity does not explain anything, a possible response is to grant a sense in which that is correct but deny that this sense of explanatory failure undermines the coherence of the distinction or show that there are no instances of "analytic." If being analytic within some linguistic practice is elucidated as "a statement understood as an empirically indefeasible stipulation," then analyticity is in some respects similar to chess bishops, stop signs, and many other classificatory concepts that are plausibly taken to be "intention dependent" in a broad sense or sometimes as "social constructions." That some mini sculpture is a chess bishop (as used on some occasion) is arguably an intention-dependent fact. Noting that the item is a chess bishop can help a bystander understand why the players are moving it only diagonally. In another sense, Harman is correct, though, in that describing something as a chess bishop does not explain anything not explained by appeal to the relevant intentions constituting the item's being (taken to be) a chess bishop. If we ask why it tends only to move diagonally on a board, this fact is explained, if at all, by appeal to the beliefs and intentions of the game players, including their belief that the piece is a chess bishop, along with associated dispositions of use and related intentional states. If being analytic is in this respect like being a chess

bishop or a stop sign, then it is importantly disanalogous to witch-hood, which is commonly understood to involve special magical powers, for example, rather than constituted by broadly social factors.

Metaphysical Versus Epistemic Accounts

More recently, Timothy Williamson and some others have distinguished between *epistemic* conceptions of analyticity and *metaphysical* conceptions. According to Williamson, both conceptions fail as accounts of a notion of analyticity that would be philosophically interesting or important. Williamson understands epistemic conceptions to be conceptions that take analytic truths to be those that any competent speaker of the language must accept. Williamson then argues that any statement can be rejected by a speaker who is linguistically competent. For example, someone might have strange conspiratorial views according to which "bachelors are unmarried men" was initially spread as part of a mind control project and is really false, that bachelors are actually twice married women. It might be unclear whether such a person is competent in English. If we tell him, "No, bachelors really are unmarried men," he might respond, "Of course, I know that's what all of you drones have been brainwashed to believe. It's false, though." It is difficult to explain why such a person should count as incompetent in English, as opposed to merely having bizarre views. According to Williamson, this sort of consideration can be generalized to any other statement.

As to "metaphysical" conceptions, Williamson takes these to involve a special type of truth property had by analytic statements. He argues that there is no good reason to think that there is more than one type of truth.

Analyticity on the Contemporary Scene

Quine was thought by many to have once and for all undermined the distinction between analytic and synthetic statements. Developments since Kripke (in the 1970s) had made some philosophers more optimistic that analyticity could be resuscitated, while for other philosophers they have seemed to sound the death knell for reasons other than those provided by Quine. Gillian Russell argues in her book *Truth in Virtue of Meaning* that there are indeed analytic statements. Her paradigm example of such a statement is "I am here now." Such statements have the distinctive characteristic, Russell argues, that any token utterance of that sentence in a context will be evaluated as true in that context, in virtue of reference-determining facts about the words contained within, along with the sentence's structure. According to Russell, this is a semantic profile for sentences that does not have anything like the epistemic significance that logical empiricists took it to have. In that respect, her view takes Quine to be right about the epistemic bankruptcy of analyticity but wrong about the semantic bankruptcy. Gillian Russell also investigates a question suppressed in most philosophical discussions (including the present one), namely, whether analyticity is in the first instance a feature of sentences, propositions, statements, utterances, or thoughts. Others influenced by possible-worlds semantics and metaphysics have come to think that there is no room for analyticity for reasons other than those given by Quine. *Semantic externalists* think that semantic relations (including truth) are "external" to minds in ways that can make them epistemically inaccessible, in a sense incompatible with a more "internalist" picture within which analyticity was introduced. Analyticity thus remains under attack along two very different fronts: the radically empiricist Quinean front and the metaphysical possible-worlds front.

Philosophers continue to draw a distinction between *conceptual truths,* stipulations, and definitions, on the one hand, and *substantive* claims, on the other. Relatedly, metaphysicians often take it that there are metaphysically "thick" ontological claims versus merely "thin" ones that seem to be akin to analytic statements of existence. Mathematics and its relation to empirical science remains ill understood, and some optimists retain the hope that appeal to analyticity in some form may illuminate that domain better than contemporary Neo-Quinean empiricist accounts. Debates in *meta-ontology,* such as those concerning the genuine existence of mereological sums, may also be well served by a new look at the analytic/synthetic distinction. The ringing of the death knell of the distinction by its many opponents may yet prove premature.

Cory Juhl

See also A Priori and A Posteriori; Duhem-Quine Thesis and the Social Sciences; Empiricism; Logical Positivism/Logical Empiricism; Metaphysics and Science; Verificationism

Further Readings

Ayer, A. J. (1946). *Language, truth and logic*. New York, NY: Dover.

Carnap, R. (1956). *Meaning and necessity: A study in semantics and modal logic* (Enlarged ed.). Chicago, IL: University of Chicago Press.

Friedman, M. (1997). Philosophical naturalism. *Proceedings and Addresses of the American Philosophical Association, 71*(2), 5–21.

Grice, H. P., & Strawson, P. F. (1956). In defense of a dogma. *Philosophical Review, 65,* 141–158.

Harman, G. (1967). Quine on meaning and existence, I. *Review of Metaphysics, 21,* 124–151.

Juhl, C., & Loomis, E. (2010). *Analyticity.* New York, NY: Routledge.

Kant, I. (1965). *Critique of pure reason* (N. K. Smith, Trans.). New York, NY: St. Martin's Press.

Russell, G. (2009). *Truth in virtue of meaning*. New York, NY: Oxford.

Quine, W. V. O. (1953). Two dogmas of empiricism. In W. V. O. Quine (Ed.), *From a logical point of view* (pp. 20–46). Cambridge, MA: Harvard University Press.

Williamson, T. (2008). *The philosophy of philosophy.* Oxford, England: Wiley-Blackwell.

Androcentrism and the Philosophy of Science

The term *androcentrism* derives from the Greek *andros* genitive of *aner* (*ἀνήρ*), for "man" or "male." This word and its variant *androcentric* are used to describe perspectives, research, theories, and methods that emphasize the behavior, activities, or experiences of males, and minimize or exclude those of females, although the research or theory might be understood to be gender neutral. Since the mid-1970s, feminist scientists and philosophers of science have argued that androcentrism is common in the methods, research questions, and hypotheses advanced in the biological, biobehavioral, and social sciences and that its presence is consequential. This entry focuses on androcentrism in the social sciences. It begins with arguments for its role advanced by feminist scientists and philosophers of science. The next sections summarize, respectively, what feminist philosophers of science argue to be some general implications of androcentrism for the philosophy of science and quite different views of the implications held by more traditional philosophers.

Androcentrism, as opposed to sexism (the view that one sex is superior), is a subtle phenomenon, one that scientists might not recognize as informing their questions, emphases, or theories. Feminist scientists and philosophers point to several dichotomies—in this case, classifications understood to divide opposed sets of behaviors, characteristics, or spheres along the lines of gender—as influencing 19th- and 20th-century social science. The dichotomies date back to the works of Plato and Aristotle.

"Gendered" Dichotomies

Examples include reason/emotion, production/reproduction, culture/nature, public/private, objectivity/subjectivity, and universality/particularity. In each, feminists argue, men are typically associated with the first category, women are associated with the second, and the characteristic, entity, or activity associated with men is taken to be worthy of study and superior to that to which it is opposed.

Some examples are in order. In psychology, Jean Piaget and Lawrence Kohlberg each developed an influential theory proposing that moral and cognitive development occur in stages. When it was observed that girls and women failed to reach the "higher stages" proposed in the theories, many assumed that this reflected a flaw in women's development rather than in the theory. Often the flaw was taken to involve one or more of the dichotomies noted: Girls and women were taken to be more subjective, and/or more emotional, and/or more concerned with particulars as opposed to universal principles than their male counterparts.

In anthropology, archaeology, economics, history, and political science, feminist scientists argued that the emphasis on what was perceived to be cultural, productive, and/or "public" effectively erased those aspects of social life associated with women (i.e., assumed to be natural, involving reproduction and/or the so-called private sphere). For example, cross-cultural studies undertaken in anthropology until the 1970s emphasized activities and behaviors associated with men (e.g., hunting rather than gathering in hunter-gatherer groups, although it would be discovered that gathering provides the majority of sustenance for such groups). Research in economics and political theory often assumed the public/private and

productive/reproductive dichotomies and focused on understanding activities associated with the first of each pair. Yet, feminists argued, the unpaid labor undertaken by middle-class women in the "private" sphere of capitalist economies deeply supported the practices and institutions in the public sphere.

The Implications of Androcentrism for the Philosophy of Science

The consequences of androcentrism (if and when it has a role in science) are relatively uncontroversial. In the foregoing examples, observations and explanations were at least incomplete, if not distorted. In addition, there was a lack of attention to issues of interest to women, although they were clearly relevant to accounts of social life—including reproduction, the so-called private sphere, rape, domestic violence, and job and wage discrimination.

However, there are disagreements about the sources and implications of androcentrism. In general, feminists argue that they reveal relationships between androcentrism and traditionally male-dominated science communities and between women's entering the social sciences following the collapse of barriers that prevented their participation and the recognition of androcentrism. They also view feminist engagements as having both a critical dimension (discussed in the previous section) and a decidedly constructive dimension. Scientists who view gender as a category warranting attention in their research, feminists point out, have identified new areas of research and proposed alternatives to traditional hypotheses. Hence, they propose a complex relationship between the sociopolitical context, on the one hand, and the directions and content of science, on the other. More substantive and detailed work on these issues is undertaken in feminist epistemology and feminist philosophy of science.

For more traditional thinkers, neither feminism nor gender does, or should, play a role in science. Cases of androcentrism are cases of "bad science," not typical of science. They also maintain that the critical and constructive developments feminists cite are the product of science's self-correcting nature, not feminism. Some charge that feminist arguments citing relationships between sociopolitical contexts and the content of science entail relativism by blurring the distinctions between science and values, and/or science and politics, and/or objectivity and subjectivity.

Somewhat ironically, feminist social scientists and philosophers cite developments within the so-called mainstream philosophy of science that, as early as the 1960s, challenged these distinctions, or at least understandings of them. These include the roles that theories have in shaping observations, that values have in scientific practice, and that the social organization and norms of a science community have in determining the weight given to views dissenting or different from those that are dominant. These developments reflect a more general recognition that, to be empirically warranted, the philosophy of science must study and reflect the actual practices of specific sciences, not idealize them.

Lynn Hankinson-Nelson

See also Feminist Critiques of Social Science Applications; Feminist Epistemology; Objectivity; Observation and Theory-Ladenness; Value Neutrality in Science

Further Readings

Crasnow, S. (2006). Feminist anthropology and sociology: Issues for social science. In D. Gabbay, P. Thagard, & J. Woods (Eds.), *Handbook of the philosophy of science* (Vol. 15, pp. 827–861). Amsterdam, Netherlands: Elsevier B.V.

Gilman, C. P. (1911). *The man made world, or our androcentric culture*. London, England: Allen & Unwin.

Hankinson-Nelson, L., & Nelson, J. (Eds.). (1997). *Feminism, science, and the philosophy of science*. Dordrecht, Netherlands: Kluwer Academic.

Harding, S. (1991). *Whose science? Whose knowledge?: Thinking from women's lives*. Ithaca, NY: Cornell University Press.

Longino, H. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton, NJ: Princeton University Press.

Annales School

The *Annales* school refers to a group of historians in 20th-century France who have been described as the most influential historiographical movement of

the 20th century, not only in France but also internationally, particularly in Latin America and India. Peter Burke, in his book on the school, spoke of the group as having initiated a "historical revolution" in France. This revolution consisted in the turn from the established narrative history, focusing on politics, events, and great men, as practiced at the Sorbonne, to a problem-oriented interdisciplinary social and cultural history. Lucien Febvre (1878–1954) and Marc Bloch (1886–1944) laid the foundations of the *Annales* group with a journal, the *Annales d'histoire économique et sociale,* launched in 1929. Although an important study of the movement spoke of an "*Annales* paradigm," the *Annales* throughout its history was marked by considerable diversity, so that the term *paradigm* hardly applied. What held it, and also later what held the group connected with the journal, together was the commitment to exploring new historical interdisciplinary approaches. One can distinguish a number of stages through which the journal went between 1929 and now. But not only did the *Annales* assert an important influence on the reorientation of historical writing in France and elsewhere, it also was very much affected by new orientations internationally.

Febvre and Bloch had already gone in very different directions. Both were deeply influenced by the *Année sociologique* of Émile Durkheim (1858–1917) and the *Revue de synthèse* of Henri Berr (1863–1954), but in different ways and to different degrees. Durkheim's stress was on social and economic structures governed by an unconscious collective mentality; Berr, whose journal, founded in 1900, was in many ways a forerunner of the *Annales,* also sought a "synthesis" that broadened the historical outlook to include all aspects of society and culture, but in which the views and intentions of individuals still played an important role. This is reflected in Febvre's early work on human geography and later in his biography of Martin Luther and in his examination of François Rabelais and the problem of unbelief in the 16th century. In the latter work, the focus is not on Rabelais as an individual but on an impersonal factor, the language of the time, which structured religious thought and made unbelief impossible. By contrast, in Bloch's *Feudal Society* (1940) individuals hardly appear, but the worldview and the cultural context of the time are central.

Until the outbreak of World War II, the *Annales* were largely the work of Febvre and Bloch, who were relatively isolated in the French historical profession. Bloch was murdered in 1944 by the Nazis. Febvre succeeded after 1945 in giving the *Annales* a firm institutional basis in the newly formed Sixth Section of the *École Pratique des Hautes Etudes,* which was transformed in 1972 into the *École des Hautes Etudes en Sciences Sociales.* The title of the journal after it was revived in 1946 was changed to *Annales: Economies, Sociétés, Civilisations* to reflect its broadened perspective. The École was committed to integrating history and the social science disciplines within a comprehensive "science of man" (*science de l'homme*) that would include not only the traditional social sciences but also linguistics, semiotics, the sciences of literature, and the arts. The École became the most important center of historical and social science studies in France, and the *Maison des Sciences de l'Homme,* established in the Parisian Left Bank in 1963, became an integrated center for these interdisciplinary studies.

In the 1960s, the *Annales* followed the fascination with quantification current at the time in the social sciences. *Annales* historians increasingly wanted to be *scientists*. They often called their institutes "laboratories" and stressed that their work needed to be quantitative to be scientific. At the same time, in contrast to the *Annales* of the 1930s, which still gave attention to the problems of an industrial society globally, studies now largely avoided the contemporary world and turned to the Ancien Régime and the Middle Ages. Emile Labrousse's (1895–1986) exploration of the origins of the French Revolution emphasized economic forces and relied heavily on quantitative methods. Regional studies focused on demographic factors in premodern French societies. This reliance on the hard sciences was reflected in Emmanuel Le Roy Ladurie's (b. 1929) *History of Climate Since the Year 1000* (1967). Quantification played less of a role in the work of Fernand Braudel (1902–1985), who followed Febvre as director of the École. In his intercultural study of the Mediterranean in the Age of Philippe II, social structures to an extent determined by geographic factors occupied a central role. What mattered in history were not short-term changes but what Braudel called the *longue durée,* long-term structural aspects. The computer entered in the 1970s with the analysis by Michel Vovelle

(b. 1933) of thousands of wills from the 17th and 18th centuries to establish changing attitudes toward death as part of a process of secularization. In the 1970s and 1980s, this reliance on quantification and the search for large-scale social structures was challenged, as historical studies in general increasingly questioned social science approaches. The work of Le Roy Ladurie in this period reflects this reorientation. His *Peasants of Languedoc* (1967) applied a Neo-Malthusian model to analyze the interrelation of food prices and population pressures in the period from the Black Death in the 14th century to the 18th century, in which individuals and politics are hardly mentioned. Individuals reappear eight years later in his story of an early-14th-century southern French heretic village, *Montaillou* (1975), a foray into historical anthropology in which a set of narratives portrays men and women embedded in an age-old folk culture. An important collection of essays, *Lieux de Mémoire* (*Realms of Memory: Rethinking the French Past*), was initiated in 1984, stressing the role of memory rather than quantitative or documentary sources in reconstructing the past. In 1994, the subtitle of the journal *Economies, Sociétés, Civilisations,* which had consciously eliminated politics, was dropped and was replaced by *Histoire, Sciences Sociales*. A broadly global perspective, which included the role of politics and religion in the past and present, now occupied more space and left room for a broad diversity of the historiographical outlook. The *Annales* continues to exist, but the question nevertheless arises whether at this point one can still speak of a specifically *Annales* orientation.

Georg G. Iggers

See also Narrative in Historical Explanation; Paradigms of Social Science; Philosophy of History

Further Readings

Burguière, A. (2009). *The Annales school: An intellectual history.* Ithaca, NY: Cornell University Press.

Burke, P. (1990). *The French historical revolution: The Annales school, 1929–1989*. Cambridge, England: Polity Press.

Iggers, G. G. (2005). *Historiography in the twentieth century: From scientific objectivity to the postmodern challenge*. Middletown, CT: Wesleyan University Press.

Stoianovich, T. (1976). *French historical method: The Annales paradigm*. Ithaca, NY: Cornell University Press.

Argumentation

This entry presents the field of argumentation, the major approaches that flourished within it, and its relevance to philosophy and the social sciences.

The study of argumentation, as distinct from the study of rhetorical theory focused on literature, has been a field of its own since 1958, when two pioneer works, Chaïm Perelman and Lucie Olbrecht-Tyteca's *Treatise on Argumentation* and Stephen Toulmin's *The Uses of Argument,* were published. These two books gave rise to distinct traditions, but both challenged models of reasoning inspired by logic and by geometry or other axiomatized sciences. Both approaches regarded juridical and judicial reasoning as more relevant in general patterns of everyday reasoning. Perelman's school in Brussels triggered numerous investigations on legal argumentation. Toulmin forged a model of argumentation as elementary and simple as the paradigmatic Aristotelian syllogism but better suited for the everyday types of reasoning (e.g., universal premises are replaced by general premises or "warrants" that tolerate exceptions). This model renewed communication studies, along with rhetorical theory in the work of Wayne E. Brockriede and Douglas Ehninger. Furthermore, Toulmin's identification of field-dependent aspects as distinct from field-invariant aspects of argumentation converged with Charles A. Willard's analysis of the sociological dimension of argumentation.

Numerous logicians, more or less explicitly in reaction against these antiformalist trends, have searched for what they call "informal logic," a phrasing that is somewhat paradoxical not only by definition but also because many of these scholars have been searching for *formal*—though *not axiomatic*—ways of modeling reasoning. One of the most successful scholars working in this vein is Jean-Blaize Grize, who has explored alternative ways of formally accounting for "the natural logic" of everyday reasoning, from Gerhard Gentzen's "natural deduction" to mathematical mereology. Researchers in computer science also search for new kinds of formal logic that are able to grasp the "nonmonotonic" aspects of everyday reasoning, namely, the frequent revision of principles—or "warrants"—and conclusions during the process of reasoning itself.

Other traditions were launched independently. Charles Hamblin, in particular, relying on an

impressive history of the various theories of fallacies from Aristotle to Stuart Mill, including the specifically Indian tradition in logic, focused on the fact that most alleged fallacies are communicative and not logical mistakes. He then opened the door to a significant renewal of the theory of fallacies, best represented by Frans Van Eemeren and Rob Grootendorst's "pragma-dialectic" theory, inspired by Paul Grice's theory of speech acts. John Woods and Douglas Walton's logical analyses have taken account of this significant pragmatic recovery of the analysis of fallacies to more precisely identify the specific role of formal logic in the analysis of everyday reasoning.

The pragmatic or speech act tradition has also inspired two closely related approaches. Oswald Ducrot and Jean-Claude Anscombre have ambitiously tried to reconstruct all of language analysis by using the basic idea that every word has an inherently argumentative aspect. In this approach, each word can be described as a particular "bundle" of commonplaces or *topoi* and each language as a structure of these *topoi,* or, in other words, of the various implicit cultural principles embedded in language and able to justify specific claims. Such an analysis is not far from Toulmin's idea of field-dependent warrants, but more than all these authors, Dan Sperber—relying on anthropological investigations—has highlighted the contextual dimension of meaning. The argumentative meaning of a myth (and, more generally, of any *topos* or warrant) depends significantly on the context of its utterance. However, unlike Grice, Sperber (along with Deirdre Wilson) has argued that not only do speakers not have to be aware of the most basic pragmatic rules allegedly guiding the interpretation of statements but these rules are reducible to a single principle, the *relevance principle* ("Be relevant when you speak!"), a functional principle of the mind. Sperber recently went further in his analysis of communication within this new cognitive framework by arguing that the emergence of reason itself might have been triggered by the intention to persuade others.

The theory of logical (and not communicative) fallacies has also been renewed by cognitive psychology (Amos Tversky and Daniel Kahneman). Stuart Mill's theory of fallacies—the last one before Frege and Russell's revival of logic—had greatly influenced one of the leading fathers of both contemporary economics and sociology, Vilfredo Pareto, who devoted a significant part of the *Tratta di Sociologia Generale* to lay out a treatise on social rhetoric. This aspect of Pareto's work has been rediscovered by Raymond Boudon and emphasized by Alban Bouvier, both in the continuity of cognitive psychology. Boudon and Bouvier have connected the study of argumentation to another general issue in the social sciences, namely, the relevance and the boundaries of rational choice theory to account for the apparent irrationality of collective beliefs.

Argumentation theory is now being reconsidered by social epistemology—or the *normative* social theory of knowledge—and by cognitive sociology—or the *descriptive* social theory of knowledge—as a double-sided part of their closely related domains (see Alvin Goldman, *Knowledge in a Social World*). Social epistemology, for example, deals with questions such as this one: How *should* the members of a jury proceed in court when they have to decide which expert is right out of two experts who disagree about the judicial responsibility of a defendant and to persuade the other jurors? Cognitive sociology (pioneered by Aaron Cicourel) deals with a complementary concern: How *do* they effectively proceed? This distinction between normative and descriptive aspects—essential to a theory of logical fallacy—was explicit in Toulmin's work but less clear in Perelman's. Evidently, the interactionist conception of communicative fallacies can be easily integrated into these complementary fields, which could unify argumentation analyses.

Finally, it should be mentioned that Jürgen Habermas has emphasized the necessary ethical dimension of argumentation, beyond pragma-dialectic rules as well as epistemological norms, and argued for argumentation ethics (or "discursive ethics") and for "deliberative democracy" in politics.

Alban Bouvier

See also Cognitive Sciences; Communicative Action Theory; Dialogical Logic; Social Epistemology

Further Readings

Boudon, R. (1994). *The art of self-persuasion.* Cambridge, England: Polity Press. (Original work published 1990)

Bouvier, A. (2007). An argumentativist point of view in cognitive sociology. *European Journal of Sociological Theory, 10,* 465–480.

Cicourel, A. (1973). *Cognitive sociology: Language and meaning in social interaction.* London, England: Penguin.

Goldman, A. (1999). *Knowledge in a social world*. Oxford, England: Oxford University Press.
Habermas, J. (1990). Discourse ethics: Notes on a program of philosophical justification. In J. Habermas (Ed.), *Moral consciousness and communicative action* (pp. 43–115). Cambridge: MIT Press.
Hamblin, C. L. (1970). *Fallacies*. London, England: Methuen.
Mill, J. S. (1963–1991). *The collected works of John Stuart Mill: Vol. 7. A system of logic, ratiocinative and inductive* (J. M. Robson, Ed.). Toronto, Ontario, Canada: University of Toronto Press. (Original work published 1843)
Pareto, V. (1935). *The mind and society*. New York, NY: Harcourt & Brace. (Original work published 1916)
Perelman, C. H., & Olbrecht-Tyteca, L. (1969). *The new rhetoric: A treatise on argumentation*. Notre Dame, IN: University of Notre Dame Press. (Original work published 1958)
Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford, England: Blackwell.
Toulmin, S. (1958). *The uses of argument*. Cambridge, England: Cambridge University Press.
Tversky, A., Kahneman, D., & Slovic, P. (1980). *Judgment under uncertainty: Heuristics and biases*. New York, NY: Cambridge University Press.
Van Eemeren, F., & Grootendorst, R. (2004). *A systematic theory of argumentation*. Cambridge, England: Cambridge University Press.
Willard, C. A. (1982). *Argumentation and the social grounds of knowledge*. Tuscaloosa: University of Alabama Press.
Woods, J., & Walton, D. (1982). *Argument: The logic of the fallacies*. Toronto, Ontario, Canada: McGraw-Hill.

Artificial Intelligence

The first recorded use of the term *artificial intelligence* (AI) was in the proposal for the Dartmouth Summer Research Conference on Artificial Intelligence in 1956. This conference—essentially a month-long brainstorming session—is usually considered the birthplace of modern AI.

However, AI has many elements that predate this. Indeed, if one takes the possibility of the mechanization of thought as the key philosophical notion underlying AI, then this notion can be found in early works by Chinese, Indian, and European philosophers. Modern AI frequently acknowledges its debt to earlier thinkers, although often in a highly selective manner and frequently as part of factional disputes.

There have been many debates on the precise definition of AI in the years since 1956. The closest thing to an agreed definition of AI within the field is attributed to Marvin Minsky, an instigator of the Dartmouth Conference: *Artificial intelligence is the science of making machines do things that would require intelligence if done by men.*

This definition may be taken as a starting point and expanded. Even if we allow no sexist implication, many researchers would not wish to be restricted solely to human intelligence. Biological research on animal behavior during the 21st century strongly suggests that intelligence has evolved more than once on Earth. Many AI researchers are influenced by these discoveries in biology and might therefore prefer a definition that included things that require intelligence *if done by animals.*

It is important to distinguish the term *artificial intelligence* from the endeavor that is Artificial Intelligence. This is because much of AI research, AI technology, and AI theory may use different names. For commercial or political reasons, the term *artificial intelligence* has sometimes been thought to convey a negative image, so the endeavor has often been rebranded or referred to solely under the name of a component technology, such as machine learning, artificial life, robotics, or advanced computer science.

The endeavor that is AI has turned out to be as much engineering as science, if not more. Many of the technological successes of AI have come from building systems that solve specific problems, rather than from first developing an understanding of the scientific principles—particularly in the many areas where human intelligence is poorly or not understood. The attempt to reproduce human intelligent behavior in machines has often revealed a partial or complete ignorance of the way in which humans achieve such behavior. In this respect, AI has prompted and pushed the science of human intelligence rather than followed its development.

A better, expanded definition would therefore be as follows: AI is the study of intelligent behavior (in humans, animals, and machines) and the attempt to find ways in which such behavior could be engineered

in any type of artifact. It is important to include "any type of artifact" because some intelligent behavior is produced by the right sort of computer software, some by the physically right sort of robot, and some by a combination of the two. Indeed, much successful AI has only a virtual existence as portions of computer code within larger systems.

AI as a Science

The science of AI is essentially interdisciplinary—involving at least computing, neuroscience, biology, psychology, mathematics, and philosophy. It is closely related to the *cognitive sciences,* though the terms are not used interchangeably. As a field of study, AI is often involved in disputes over methods and areas with these more established areas of science. This can be aggravated by the tendency to use the term *artificial intelligence* only for the more experimental or speculative portions of the research, for reasons alluded to in the previous section. The tendency to give AI successes a different name also applies to AI as a science.

The science of AI can be seen as having progressed through three distinct phases. These phases are not properly to be called *paradigms* in Thomas Kuhn's use of the term, since they are compatible and modern. AI technology often utilizes an eclectic mix of ideas from all three phases.

The first phase is known as GOFAI (good old-fashioned AI). During this phase, the most important route to intelligent behavior was considered to be an appropriate algorithm. That is usually a clever programming technique capable of deducing or guessing the best solution to a problem. This is the phase of AI that most resembles conventional computing. It is important to stress that this was, and is, more than mere computation in that AI programs usually operate in areas where informed guesswork is necessary.

A good example would be game-playing AI, which, among other successes, has produced chess-playing programs capable of beating even the best human players. It is often believed that these programs play chess by calculating all possible moves. This is not the case. It is mathematically impossible to calculate all the moves available in a chess game. A chess-playing program must make informed guesses as to which moves are more likely to lead to a winning position. These informed guesses are known as *heuristics,* and they are refined by letting the program play games (usually against itself).

The second phase stemmed from an increasing realization that GOFAI methods, however successful, were very unlike the operation of natural brains, leading some researchers to explore very different approaches. This second phase was characterized by attempts to produce software that was more human or animal-like in its operation. This focus of interest used names such as *parallel distributed processing* or *connectionism.*

A crucial difference from GOFAI is that connectionist methods do not depend on sequential processing of information and rules. Instead, multiple processors make aggregate decisions in an imitation of the neurons in a human or animal brain. It is important to observe that this is a fairly loose imitation, as the operation of neurons is as yet incompletely understood.

Connectionist methods can often produce behavior that is much more brainlike than that produced by GOFAI methods. Features such as partial remembering or the ability to synthesize the missing portions of incomplete data also make these techniques useful in real-world applications.

The third phase was characterized by a much greater emphasis on the building of physical robots. This is usually known as *situated robotics* or *artificial life*. The change of emphasis in the third phase is an explicit rejection of producing intelligent behavior purely in simulation or from "disembodied software." It therefore focuses on *embodiment* as an essential component of intelligent behavior. This third phase has produced much more useful robots than the previous two phases—aided by concurrent developments in mechanical engineering and battery technology.

It is often argued that psychology has learned much from AI over the past 50 years. One way in which this can occur is in the building of computer models of human thought processes. The level of understanding of the mental processes required in order to do this has imposed a useful attention to detail, which would otherwise be lacking in psychological theories.

Many psychologists acknowledge the direct influence of AI. Even among those who claim to be not directly influenced by AI theories, it is generally

agreed that the development of AI has helped inspire the emergence of more scientific approaches to psychology.

AI as Engineering

No one who uses modern technology can have failed to notice that many devices are getting smarter. Much of AI technology is computer software that produces intelligent behavior in a particular application. This can be seamlessly and often invisibly integrated into computer systems that control larger systems—and human behavior—in the real world.

A good example would be the now common use of AI in traffic management systems. Rather than having human operators make a vast number of complex decisions in order to optimize traffic flows and minimize pollution levels from vehicle emissions, road systems are being increasingly turned over to automatic control. Automatic control is much faster and more reliable than human operators. The AI components in such systems are pieces of computer software that solve complex problems in real time. Without close examination, one would not know that this AI technology was there. Note that such systems also regulate human behavior in that drivers are legally required to comply with their instructions via traffic signals and variable speed limits.

This example illustrates many of the key features of AI as engineering. Without specialist inside knowledge, the public are often not aware of its existence—especially its presence in social life, controlling and enabling social action. In many cases, the developers, owners, and users of automated systems do not want the label "artificial intelligence" applied to their system. Such a system will be more acceptable to the public if it is described by what it does, without any mention of AI technology.

AI technology is involved in the monitoring of financial transactions and in detecting credit card fraud in real time. AI technology is essential for the routing of mobile phone calls. Many online advice and help systems use AI technology in a perhaps more obvious way. Car, aircraft, and railway engines are often monitored by AI programs. These applications are just as likely to use connectionist approaches or a mixture of GOFAI and connectionist methods.

Internet search engines are examples of applied AI. Similarly, the attempt to get intelligent behavior from robots, although at an early stage, has resulted in a number of toys that can learn and in military robots that possess a useful degree of autonomy. It is difficult, if not impossible, to exhaustively list the applications of AI as engineering, simply because it pervades so much modern technology.

The Myth of AI Failure

Despite the fact that AI technology is all around us, a myth has developed both inside and outside AI that it is a failed project. This myth is completely false—AI has been remarkably successful as both science and engineering. There are, however, many reasons for the emergence of such a myth.

An important reason within the field of AI is the tendency to promise major breakthroughs on ridiculously short timescales. This has led to disillusionment in funding bodies and to a number of so-called AI winters, during which it has been difficult to continue research.

Also important in the emergence of the myth has been the often highly skeptical response of some philosophers. Many philosophers have been unimpressed by the engineering successes of AI. They claim that these achievements do not in any way compose "real intelligence." The philosophical arguments might be characterized as whether the word *artificial* in "artificial intelligence" has the same force as in artificial light, which is "real light," or as in artificial flowers, which are but a crude imitation and lack most of the properties of real flowers.

Philosophical Responses to AI

The Turing Test

The expression *Turing test* is applied to a game introduced in an article, "Computing Machinery and Intelligence," published in 1950 in *Mind*—one of the longest-established British philosophy journals. It was written in 1948 by Alan Turing, a mathematical genius who had spent the war years of 1939 to 1945 at Bletchley Park engaged in secretly breaking the German Enigma code. By 1948, Turing had moved to Manchester, where an electronic computer—arguably the first modern computer—had been constructed. At the time when he was writing "Computing Machinery and Intelligence," Turing was also engaged in writing a program code for this machine. Importantly, this article is therefore

a philosophical article written by a mathematician turned code breaker turned computer programmer.

In "Computing Machinery and Intelligence," Turing says he wishes to discuss the question "Can machines think?" However, since this question is too vague, he proposes replacing it with a game. This game he called "the imitation game." It involves three people in separate rooms. They can communicate only by typing messages to each other. In the original version, there are a man, a woman, and an interrogator whose gender is unimportant. The interrogator, as the name suggests, can ask any question of the other two participants. The objective of the game is for both the man and the woman to convince the interrogator that they are the woman. The woman will be answering truthfully, and the man will be typing things like "Don't listen to him, I'm the woman."

Now what would we say, asks Turing, if the role of the man in this game were to be successfully played by a machine? That is if, after 5 minutes of questions, the average interrogator would not be able to recognize that he or she was communicating with a machine at least 30% of the time.

If the time comes when machines can achieve this level of success in the imitation game, then, so Turing claims, people will be happy to describe such machines as thinking machines. The imitation game has become known as the *Turing test,* and many people think that achieving Turing's level of success in the imitation game is a test of whether or not AI has finally been achieved.

Turing thought that it was a matter of *when,* not *if,* we would make such machines. He confidently predicted that by the year 2000, digital computers would be able to achieve this level of success in the imitation game. This achievement would change public attitudes so that it would become normal to talk of "thinking machines." However, no computer is anywhere near good enough to succeed in the imitation game in the foreseeable future. One important problem in building a machine capable of winning the imitation game is that the interrogator may ask absolutely any question and that humans seem to be very good at detecting implausible answers.

Unfortunately, using the imitation game as a test of AI is highly misleading. It has often distracted AI researchers into human imitation and methods of deceiving people. Such imitation is not central to AI as either science or engineering. Nonetheless, Turing's article has become central to the philosophy of AI. A key philosophical principle introduced in this article is that of *functional equivalence*. If a computer produces responses that are indistinguishable from those of a human, then in everyday speech, we would be inclined to say that it can think.

The Chinese Room

A much-cited objection to the claim that computers could ever think is attributed to the philosopher John Searle and has come to be known as the "Chinese room" argument, after the thought experiment he proposes. Despite many criticisms, the Chinese room argument continues to be discussed after more than 30 years.

Searle purports to show that an appropriately programmed computer cannot think. This may turn out to be true as a matter of science—though it seems unlikely on present evidence. Searle does not wait for science to resolve the issue: He claims to be able to show by argument alone the impossibility of a machine thinking as a result of following a program.

He asks us to consider a room that gives reliable and plausible answers to questions in Chinese. What is interesting about Searle's room is its internal working. Inside the room is a comprehensive set of instructions (written in English) detailing how to respond to any input with Chinese symbols. He (Professor Searle) sits inside the room following the instructions and, thereby, can respond appropriately to any input question, though he does so without ever knowing what any of the symbols mean or what the questions are about.

Searle does not mention Turing's imitation game explicitly but clearly thinks he has performed a *reductio* on Turing. Even the capability of answering any possible question does not entail that the machine is thinking. It may be simply *appearing to think* by following a program.

Many people in AI have responded to this thought experiment over the years, but Searle remains obdurate, dismissing them as "all wrong." Some of those responding think that the "Chinese Room" needs to be put into a real robot in the real world—as in the third phase of AI.

The most frequently made response is called by Searle as "the Systems Response." The main claim

of this response is that although the Searle in the room does not understand Chinese, the system as a whole does understand Chinese. The systems reply is obviously wrong also, says Searle, because just as there is no understanding in the Searle in the room, there is no understanding anywhere else in the system either. The book of instructions is just a book, and the room is just a room. Nothing in the system understands Chinese any more than Searle does, so it's pointless to look for understanding in the system. It is arguable that Searle's obduracy is mistaken, because it has often been observed (at least as early as in Leibnitz) that no *part* of a thinking system has to be able to think. Understanding therefore is probably not to be sought in any *part* of the Chinese Room.

The Chinese Room argument has proved enduring. Some of the reasons for this lie in its appeal to our intuitions. We do not feel as if our thoughts are, or even resemble, following a program. Nor does it seem plausible, at first glance, that the richness of mental lives could ever be reproduced by a machine that follows instructions. The notion of a thinking machine remains controversial, despite the success of AI.

The Ethical Implications of AI

Philosophical responses to AI have concentrated on the conceptual issue of whether or not AI can be said to really think and whether or not the idea of making a thinking machine is conceptually coherent. For this reason, there has been much less attention paid to the ethical implications of the changes being brought about by the actual use of AI. The invisibility of successful AI technology also contributes to a lack of awareness of the social and ethical issues around AI.

There is now a growing interest in some of the pressing ethical issues being thrown up by AI. Consider, for example, the use of AI in medicine. Modern knowledge-based systems have been shown to be superior to human doctors in certain specialized areas. There is also evidence that psychotherapy is sometimes more effective if performed by a computer than if performed by a human therapist. It would be unethical to resist the introduction of AI in these areas. However, there are no clear ethical principles established for the design and use of such systems. As yet, there have been no criteria agreed on as to who should be held responsible when things go wrong.

There is now international concern about the use of autonomous military robots, with some authorities calling for a moratorium on their deployment. The use of robots in caring applications—for example, in child care—and in "smart homes" for elderly people also raises ethical issues that remain unresolved.

Blay Whitby

See also Agent-Based Modeling and Simulation in the Social Sciences; Classical Computationalism, Connectionism, and Computational Neuroscience; Coalition Logic; Cognitive Sciences; Embodied Cognition; Human–Machine Interaction; Machine Consciousness and Autonomous Agents; Mind–Body Relation; Multi-Agent Modeling

Further Readings

Boden, M. A. (Ed.). (1996). *Artificial intelligence: Handbook of perception and cognition.* London, England: Academic Press.

Boden, M. A. (2006). *Mind as machine: A history of cognitive science* (Vols. 1 & 2). Oxford, England: Oxford University Press.

Dreyfus, H. (1992). *What computers still can't do: A critique of artificial reason.* Cambridge: MIT Press. (Original work published 1972)

Ekbia, H. R. (2008). *Artificial dreams: The quest for non-biological intelligence.* Cambridge, England: Cambridge University Press.

Haikonen, P. O. (2007). *Robot brains: Circuits and systems for conscious machines.* Chichester, England: Wiley.

Husbands, P., Holland, O., & Wheeler, M. (Eds.). (2008). *The mechanical mind in history.* Cambridge: MIT Press.

Minsky, M. L. (Ed.). (1968). *Semantic information processing.* Cambridge: MIT Press.

Nilsson, N. (2010). *The quest for artificial intelligence: A history of ideas and achievements.* Cambridge, England: Cambridge University Press.

Searle, J. (1991). *Minds, brains and programs.* London, England: Penguin.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59,* 433–460.

Wheeler, M. (2005). *Reconstructing the cognitive world.* Cambridge: MIT Press.

Whitby, B. (1996). *Reflections on artificial intelligence.* Exeter, England: Intellect Books.

Whitby, B. (2009). *Artificial intelligence: A beginner's guide.* New York, NY: Rosen.

Austrian Economics

The Austrian school of economics owes its name to the German historical school critics of the works of the Viennese economist Carl Menger in the late 19th century. Ironically, Menger had thought his work was a contribution to the German-language scientific literature in economics and, in fact, dedicated his *Principles of Economics* to William Roscher, leader of the older German historical school. He even concluded his preface by stating that he hoped his work would be viewed as building on the developments in political economy among German scientists and thus welcomed as a friendly greeting from a collaborator in Austria. Menger did not offer this dedication with any cynical intent but in sincere appreciation.

But Menger's was a theoretical contribution, and in particular, it concerned the deductive nature of pure theory, whereas the members of the younger German historical school had thought themselves free of abstract theoretical systems of thought, which only hindered progress in the study of humankind and, more to the point, progressive reforms in public policy. Menger's defense of the necessity of a theoretical framework for doing empirical social science sparked off the *Methodenstreit* in German-language economics. It was during this dispute that the economic ideas of Carl Menger—methodological individualism, subjective utility theory, marginal analysis, dynamic competition, relative prices, and especially the defense of the deductive method—were dubbed "Austrian." The name "Austrian school of economics" has stuck ever since.

Origin and Development of the Austrian School of Economics

The main contributors to the Austrian school of economics taught or studied at the University of Vienna from the late 19th century until the upheavals in Europe caused by Hitler in the 1930s. The list of economists associated with the Austrian school during this period is very impressive and includes many that are today considered as some of the leading economists worldwide during their time: Carl Menger, Eugen von Böhm-Bawerk, Friedrich Wieser, Joseph Schumpeter, Hans Mayer, Ludwig von Mises, Fritz Machlup, Oskar Morgenstern, Gottfried Haberler, and F. A. Hayek.

Schumpeter was the first to move outside Vienna, and he also sought to distance himself from his educational pedigree. But Schumpeter's core ideas about entrepreneurship, dynamic competition, and creative destruction were intimately connected to the teachings of his Viennese mentors. Machlup, Morgenstern, and Haberler all migrated to the United States and established teaching careers at Johns Hopkins, Princeton, and Harvard in the 1940s. But it was Hayek's move to the London School of Economics in the 1930s and the central role he played there in both the debate with John Maynard Keynes and that with Oscar Lange over socialism that began to spread the teachings of the Austrian school of economics throughout the elite of professional economists in the English language. Von Mises was the mentor for Machlup, Morgenstern, Haberler, and Hayek back in Vienna, and they all worked on various aspects of von Mises's ideas. But von Mises's work was not widely available in English until the 1940s, after he migrated to the United States and started teaching at New York University.

The Modern Austrian School

The modern Austrian school emerged in the 1950s and was centered around von Mises at New York University. When Hayek moved to the University of Chicago in 1950, he was not located in the economics department but in the Committee on Social Thought, and his work during that period was focused on methodology, intellectual history, and political and legal theory and history. The modern Austrian school of economics took its inspiration from two books written in the late 1940s by von Mises and Hayek—*Human Action,* by von Mises, published by Yale University Press in 1949, and *Individualism and Economic Order,* by Hayek, published by University of Chicago Press in 1948. It was these presentations of the method and methodology of economics that would inspire the work of a new generation of graduate students in economics, who would pick up the label "Austrian economics" even though they had never taken a class at the University of Vienna.

The two leading figures of the modern Austrian school were Murray Rothbard and Israel Kirzner, who while earning their PhDs in the 1950s would make significant contributions to economic science

through the 1960s and 1970s. Rothbard, for example, published his own comprehensive treatise of economics in 1962, *Man, Economy and State,* while Kirzner published more focused monographs, such as *The Economic Point of View* in 1960, *Market Theory and the Price System* in 1963, and *An Essay on Capital* in 1966. Rothbard also branched out and wrote an economic history of the Great Depression—*America's Great Depression,* published in 1963—which countered the conventional wisdom that capitalism is inherently unstable and argued instead that the Great Depression was a consequence of government intervention and policy errors. Rothbard also wrote political theory tracts, most notably *For a New Liberty* (1973) and *The Ethics of Liberty* (1981). Kirzner, on the other hand, was more narrowly focused on economic theory, and his most famous work, *Competition and Entrepreneurship,* was published in 1973. Through these works and, of course, the still looming influence of von Mises and Hayek, a group of students began to express an interest in studying Austrian economics in depth and contributing to the scientific tradition. Rotbhard and Kirzner organized a conference in the summer of 1974 with the Institute for Humane Studies.

Recent Developments

Then in the fall of 1974, Hayek was awarded the Nobel Prize in Economic Science, and new attention was paid to both the historical contributions of von Mises and Hayek and the group of new scholars influenced by Rothbard and Kirzner, who were joining the ranks of professional economists through the 1970s and 1980s. In addition to this renewed scientific interest, events and the tide of public opinion had turned in a way that was more receptive to the basic laissez-faire message of classical economics and the Austrian school. The stagflation of the 1970s and the collapse of communism in the late 1980s brought new adherents to the Austrian school of economics and its critique of Keynesianism and socialism. Works such as Hayek's *The Road to Serfdom* and *The Constitution of Liberty* were invoked by both Ronald Reagan and Margaret Thatcher, and then by the postcommunist leaders in East and Central Europe and the former Soviet Union. PhD programs were established at Auburn University and George Mason University, and of course, New York University was still the center of activity. In addition, professional scientific journals were established as well as book series. A professional scientific society was established—Society for the Development of Austrian Economics—which hosts an annual conference with well over 100 participants from throughout the world.

And since 2008, the business cycle theory of the Austrian school of economics has often been invoked by critics of government policies of easy credit during the housing boom and the bailouts of the big banks, as the major alternative to the more politically acceptable Keynesian demand management theory. Major policy intellectuals have been quoted as saying that they read von Mises at the beach—a rather absurd claim since *Human Action* runs close to 1,000 pages and is a rather densely argued philosophical as well as technical economics tract.

Epistemological Issues

The Austrian school of economics has evolved significantly since the founding period at the University of Vienna, but the basic building blocks of methodological individualism, methodological subjectivism, and market process theory—all pursued following the deductive method—remain. Substantively, you can divide the contributions of the Austrian-school economists into three categories: methodological, analytical, and ideological.

Methodologically, the Austrian school stands in opposition to both the formalist and the empiricist transformation of economics that took place in the mid 20th century. Austrian economists, because they tend to focus their intellectual attention on phenomena that defy strict formalism, will rely on verbal reasoning rather than mathematical. One way to summarize this position is that for the standard mathematical models to hold, the phenomena being studied must be assumed to be smooth, continuous, and twice differentiable, in order to lend themselves to the conventional mathematical techniques employed by economists. But according to the Austrian economists, the world consists of behavior that is "lumpy" and "discontinuous." In that world, the most logically rigorous approach may defy the formalism of the method of simultaneous equations.

Analytically, the Austrian school stands in opposition to the equilibrium-always textbook presentations of general competitive equilibrium or partial

equilibrium comparative statics. Instead, the focus is on processes of adaptation due to constantly changing circumstances. It is this focus on change and process that explains the preoccupation with entrepreneurship among Austrian economists—as the entrepreneur is by definition the agent of change in an economic system and also the agent of adjustment to those changed circumstances.

Ideologically, the Austrian school is the most consistent within the modern profession of economics in adhering to the classical liberal political economy of limited government. The main lesson of the teachings of the Austrian school, from Menger to Hayek to Kirzner, is that the market economy within the framework of secure private property rights, freedom on contract, sound money, fiscal responsibility, and free trade is not only self-correcting but also the source of generalized prosperity and improvements in the well-being of mankind.

The Austrian school of economics continues to exist both as a scientific tradition within professional economics and as an iconic rallying call for free-market activism in the political arena. The two different strands do not always exist without conflict. But from Menger onward, the Austrian economists have been committed to the proposition that policy relevance is a virtue and not a vice in economic discussions. On the other hand, the Austrian economists are also mindful of Claude Frédéric Bastiat's warning that "the worst thing that can happen to a good cause is not to be artfully criticized but to be ineptly defended."

Peter J.Boettke

See also Capitalism; Individualism, Methodological; *Methodenstreit;* Philosophy of Economics, History of; Weber and Social Science: Methodological Precepts

Further Readings

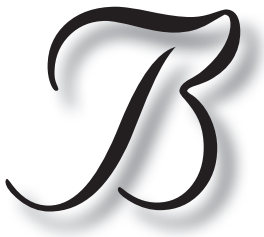
Boettke, P. (Ed.). (1994). *The Elgar companion to Austrian economics*. Cheltenham, England: Edward Elgar.

Boettke, P. (Ed.). (2010). *The handbook of contemporary Austrian economics*. Cheltenham, England: Edward Elgar.

Hayek, F. A. (1948). *Individualism and economic order.* Chicago, IL: University of Chicago Press.

Menger, C. (1981). *Principles of political economy.* New York: New York University Press. (Original work published 1871)

von Mises, L. (1949). *Human action: A treatise on economics*. New Haven, CT: Yale University Press.



Bargaining Theory

Bargaining is a universal social phenomenon. It takes place whenever humans are engaged in "antagonistic cooperation" and have to negotiate agreements on matters such as prices, wages, and regulations concerning personal, group, and international relations. Bargaining parties have partly opposing interests and need to negotiate how to compromise them to the advantage of all concerned. Bargaining theories characterize potential agreements descriptively and normatively. When and why cooperation will break down and what will, or should, be the shares after it stops in (dis)agreement are particularly relevant issues.

This entry focuses on "normative" characterizations of rational bargaining, analyzes its essential features, and points out its ethical implications. It also raises the question as to what extent the ideal theory coincides with real-world situations.

As a special case that can be generalized to $n \geq 2$ individuals, imagine two individuals, A and B, negotiating over a set of feasible outcomes. If they fail to agree, the outcome will be the status quo that each can realize without the agreement or cooperation of the other. Let $u = (u_A, u_B)$ be the pair of "cardinal" utility functions representing the preferences of A and B, respectively. Let S be the set of all "utility pairs" associated with those states that can in fact be realized. Let $d = (d_A, d_B) \in S$ be the utility pair, also called disagreement point, associated with the status quo that emerges if no agreement is reached. Only if some state of affairs $(s_A, s_B) \in S$ exist such that $s_A > d_A$ and $s_B > d_B$, there is a "bargaining incentive" for both individuals, A and B.

If S is convex and compact, the pair (S, d) is called a *bargaining situation.* It represents the underlying bargaining problem in a two-dimensional space of von Neumann–Morgenstern, "vNM," utility functions. John Nash axiomatically characterized a unique *solution function* f assigning a solution $f(S, d)$ to all situations (S, d) in the relevant utility space. For each situation (S, d), it selects the maximum of the function $(s_A - d_A)(s_B - d_B)$ within the individually rational subset S of all utility pairs such that all individuals at least realize their disagreement levels.

Nash assumes that the preferences of both actors should fulfill the vNM axioms for the existence of a class of utility functions that represent preferences. Propositions formulated relying on any such function u must remain true for $a \cdot u + b$ with $a > 0$, as the axioms do not allow us to discriminate between functions within the class. Nash's first additional axiom—beyond those guaranteeing the existence of a vNM utility representation of preferences—requires that the function evaluating the joint results should pick the same result for any positive linear transformation of each individual's utilities (at the same time avoiding interpersonal comparisons of utility in characterizing the result). The second axiom requires for a symmetric (S, d) that the solution $f(S, d)$ have identical utility components. The third axiom demands weak Pareto optimality of the solution of each situation. For each (S, d), there is no feasible point in S that would make everybody strictly better off than the one selected as a solution of (S, d). Fourth, the requirement of independence of

irrelevant alternatives means that, for a given status quo *d,* if the solution of a larger problem *S* is also part of a smaller set *S'*, then what is the agreement in the larger one should be chosen as the agreement in the smaller set as well.

That the loss of alternatives that are good for some person will not change the outcome as long as the former agreement is still available does not seem as plausible as the other axioms. In parallel to Nash, Howard Raiffa was working on a proportional solution concept that was not independent of "forgone" alternatives. Some 20 years later, Ehud Kalai and Meir Smorodinsky introduced the "ideal point," the pair of maximal utility values for each person in the individually rational subset of *S*. Substituting Nash's fourth axiom with their own, they compare situations with identical status quos and identical ideal points and require a monotonic reaction of agreement outcomes on enlargements of the feasible set. For the two-person case, they uniquely characterized a solution function on *S*. The selected solution point on the boundary of *S* has the property of assigning equal relative gains compared with the maximal gain between the status quo and the ideal point to both players, and these relative gains are maximal in *S*.

There is a wealth of spin-offs of the original approach. For instance, John Harsanyi has shown that in Frederik Zeuthen's model of concession making in negotiations, the concession process converges toward the Nash solution. David Gauthier used a different definition of a concession, where the ideal gain between status quo and ideal point is always in the focus when sizes of concessions are compared. He argues that rational players will compare and make concessions such that the process of stepwise concessions will converge toward the Kalai–Smorodinsky solution.

An examination of the normative justification of the axioms used in characterizations of bargaining solutions has obvious relations to ethical and other normative philosophical issues. At the same time, one may ask to what extent real-world bargaining empirically coincides with the ideal theory and how well, in particular, empirical processes of stepwise concession making cohere with, say, the Zeuthen–Harsanyi approach or, alternatively, aspiration adaptation and satisficing models. The project of modeling all the minute details of bargaining explicitly in noncooperative game theory and experimentally testing behavioral assumptions against the background of normative ones is still unfinished. Perhaps most extensively, so-called ultimatum bargaining has been used to check on the validity of the classical model of rational economic man in interactive decision making in the laboratory and in the field.

Hartmut Kliemt and Marlies Ahlert

See also Cooperation/Coordination; Evolutionary Game Theory and Sociality; Game-Theoretic Modeling; Homo Economicus; Markets and Economic Theory; Pareto Optimality; Preference; Social Choice Theory; Social Contract Theories

Further Readings

Binmore, K., Osborne, M. J., & Rubinstein, A. (1992). *Noncooperative models of bargaining.* In R. J. Aumann & S. Hart (Eds.), *Handbook of game theory with economic applications* (Vol. 1, pp. 179–225). Amsterdam, Netherlands: North-Holland.

Gaertner, W., & Klemisch-Ahlert, M. (1992). *Social choice and bargaining perspectives on distributive justice.* Berlin, Germany: Springer.

Gauthier, D. P. (1986). *Morals by agreement.* Oxford, England: Oxford University Press.

Harsanyi, J. C. (1977). *Rational behavior and bargaining equilibrium in games and social situations.* Cambridge, England: Cambridge University Press.

Kalai, E., & Smorodinsky, M. (1975). Other solutions to Nash's bargaining problem. *Econometrica, 43,* 513–518.

Luce, D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey.* New York, NY: Wiley.

Nash, J. F. (1950). The bargaining problem. *Econometrica, 18,* 155–162.

Peters, H. (1992). *Axiomatic bargaining game theory.* Dordrecht, Netherlands: Kluwer Academic.

Thomson, W. (1994). *Cooperative models of bargaining.* In R. J. Aumann & S. Hart (Eds.), *Handbook of game theory with economic applications* (Vol. 2, pp. 1237–1284). Amsterdam, Netherlands: North-Holland.

Zeuthen, F. (1930). *Problems of monopoly and economic warfare.* London, England: Routledge.

Bayesianism, Recent Uses of

Bayesianism, named after Thomas Bayes (1702–1761), is a family of related approaches in epistemology, philosophy of mind, Decision Theory, and statistics. In epistemology, philosophy of mind, and Decision Theory, Bayesians use probability theory to

represent rational agents' degrees of belief. In epistemology, Bayesianism is both a movement in its own right and a methodology for approaching various traditional problems. In statistics, Bayesianism is a view about statistical inference. It has been influential in most of the sciences and the social sciences, as well as in financial calculations, where probabilistic and statistical methods are standard.

The following section of this entry discusses Bayesian approaches to epistemology and philosophy of mind; the next section canvases Bayesian approaches to statistics; and the final section briefly surveys some applications of Bayesianism in the social sciences.

Bayesianism in Epistemology and Philosophy of Mind

In epistemology and philosophy of mind, Bayesianism's central tenet is that uncertainty should be represented probabilistically. While there are some intramural disputes, most Bayesians in these fields maintain that rational beliefs come in degrees that obey probability theory, beginning with a "prior" probability function before any enquiry and updating by repeatedly applying a particular rule as evidence comes in. As such, the view is associated with the "subjective" interpretation of probability (as opposed to "objective" interpretations such as frequentism).

Bayesians claim that rational degrees of belief obey probability theory. Probability theory imposes axioms on a *probability function P:* It is nonnegative, has a maximum value of 1, and is additive ($P(X \text{ or } Y) = P(X) + P(Y)$ if X and Y cannot both occur). For example, we might represent your uncertainty about how a fair die toss will land with a probability function P that assigns 16 to each of the 6 faces landing up, for which $P(\text{even number}) = P(2 \text{ or } 4 \text{ or } 6) = 16 + 16 + 16 = 12$.

The *conditional probability of X given Y* ($P(X \mid Y)$) is given by the ratio of unconditional probabilities:

$$P(X \mid Y) = PX \text{ and } YPY, \quad \text{provided } P(Y) > 0.$$

Thus, the conditional probability that the die lands 1, *given* that it lands odd, is 1/3:

$$P(1 \mid \text{odd}) = 1/3.$$

Suppose that your initial probabilities are given by P_{initial} and that you become certain of an event E (and of nothing more). What should be your new probability function, P_{new}? Bayesians favor updating by *conditionalization:*

$$\text{Conditionalization} \quad P_{\text{new}}(X) = P_{\text{initial}}(X \mid E), \quad \text{provided } P_{\text{initial}}(E) > 0.$$

For example, if you become certain that the die landed odd, your new probability for the die having landed 1 should be 1/3.

A famous result in probability theory is *Bayes's theorem:*

$$P(H \mid E) = P(E \mid H)P(H)/P(E).$$

H is a *hypothesis* and E, some *evidence*. $P(E \mid H)$ is the *likelihood*, and $P(H)$ and $P(E)$ are the *prior* probabilities. Bayes's theorem is often a convenient way to calculate the conditional probabilities that are required by conditionalization; thus, it is usually regarded as a cornerstone of Bayesianism. We stress, however, that Bayes's theorem is not *itself* an updating rule but rather a constraint on probabilities assigned by a single probability function or by an agent at a single time.

The picture of an ideally rational agent who assigns probabilities to various states of affairs, and who conditionalizes on the evidence as it comes in, goes back to F. P. Ramsey, Leonard J. Savage, and Bruno de Finetti. They also developed *Bayesian Decision Theory.* A rational decision is an action that maximizes *expected utility,* a weighted average of the payoffs associated with the action (for each possible state of affairs), the weights provided by the corresponding probabilities. These and other Bayesians have argued that an agent who violates Bayesianism's tenets will be susceptible to making bad decisions. *Objective* Bayesians impose further constraints on the agent's prior probabilities.

Bayesianism in Statistics

Bayesianism is an important and increasingly widespread view in statistics.

Suppose that we are interested in whether, and to what extent, a particular coin is biased toward heads and that we entertain various hypotheses about its bias. Suppose that we flip the coin 10 times, observing that it lands heads 7 times. What can we say about the probabilities of the various bias hypotheses, given these data? This is a question about statistical inference.

A Bayesian statistician begins with a *prior* probability distribution, before incorporating the data in question. (Note that this probability distribution need not be interpreted as any rational agent's degrees of belief, so a Bayesian statistician need not adopt a Bayesian philosophy of mind!) The prior might be based on the shape of the coin, previous observations of similar coins, and so on. This prior assigns some probability to each hypothesis regarding the coin's bias, and a likelihood for the actual data, given any particular such hypothesis. The prior is then combined with the data to obtain a *posterior* probability distribution—a synthesis of the two. (Bayes's theorem is typically used here.) Given a particular data set, a hypothesis has high posterior probability to the extent that (1) it has high prior probability and (2) the actual data are likely according to that hypothesis. The Bayesian statistician thus arrives at posterior probabilities for the hypotheses given the observed data.

The Bayesian approach to statistics is criticized for its reliance on prior probability distributions. Without clear guidelines about how to construct priors, the criticism goes, statistical inferences are too subjective. Bayesians counter this by attempting to provide the desired guidelines, by embracing the subjective nature of statistical inference, or by using data sets so large that all reasonable priors converge to similar posteriors. The main alternative to Bayesian statistics is *classical statistics*. In recent decades, improved computation has facilitated the increasingly widespread use of Bayesian techniques.

Bayesianism in the Social Sciences

The social sciences are largely concerned with human behavior. As such, they provide natural applications for Bayesian Decision Theory. Its early roots were in economics, and it continues to be influential there. Moreover, it complements *game theory*, the theory of strategic decision making by multiple agents in situations in which their profile of payoffs depends on what all the agents do individually. While John von Neumann and Oskar Morgenstern's seminal 1944 book *Theory of Games and Economic Behavior* assumed that all the relevant agents have the same probability distribution, Johann Pfanzagl in 1967 offered a Bayesian generalization by axiomatizing subjective probability and utility in their framework. Bayesian game-theoretic modeling is similarly important in political science, for example, in institutional design, voting theory, and auction theory (e.g., modeling electoral competition as an auction).

More generally, probability and statistics play an indispensable role in many branches of science and social science. As such, Bayesian methods are applied to areas as diverse as physics, biology, law, and sociology. Bayesian methods are also used as a tool for theory choice across the sciences and social sciences.

Bayesian networks (also known as directed acyclic graph models) provide a neat and intuitive way of representing causal relationships. They have a tremendous range of applications and are of particular use in modeling complex systems such as economies and the Internet. Bayesian networks are also used to develop management systems in a wide range of public policy areas, from water resource management to nuclear proliferation.

Bayesian models are often criticized for being highly idealized. In psychology, a large literature, pioneered by authors such as Ward Edwards and Herbert Simon, has explored how fruitfully these models may be deployed to understand actual human agents. A lively debate has raged over the extent to which we fall short of Bayesian ideals—Daniel Kahneman, Amos Tversky, and Gerd Gigerenzer have been the leading protagonists. Accordingly, some Bayesians have relaxed the traditional framework—for example, allowing degrees of belief to be imprecise, rather than sharp numbers. More recently, a trend has emerged of constructing Bayesian models of learning and cognition. For example, they have been used in psychology and linguistics to model language learning.

Alan Hájek and John Cusbert

See also Decision Theory; Epistemology; Formal Epistemology; Game-Theoretic Modeling; Probability; Rational Expectations

Further Readings

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory.* Chichester, England: Wiley.

Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology.* Oxford, England: Clarendon Press.

de Finetti, B. (1974–1975). *Theory of probability* (2 vols.). New York, NY: Wiley.

Edwards, W. (1961). Behavioral decision theory. *Annual Review of Psychology, 12,* 473–498.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond heuristics and biases. In W. Stroebe & M. Hewstone (Series Eds.), *European review of social psychology* (Vol. 2, pp. 83–115). Chichester, England: Wiley.

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences, 34,* 169–188.

Pfanzagl, J. (1967). Subjective probability derived from the Morgenstern-von Neumann utility theory. In M. Shubik (Ed.), *Essays in mathematical economics in honor of Oskar Morgenstern* (pp. 237–251). Princeton, NJ: Princeton University Press.

Ramsey, F. P. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *Foundations of mathematics and other essays* (pp. 156–198). London, England: Routledge.

Savage, L. J. (1954). *The foundations of statistics.* New York, NY: Wiley.

Simon, H. A. (1982). *Models of bounded rationality.* Cambridge: MIT Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Behavioralism in Political Science

The behavioral movement, sometimes referred to as the behavioral revolution, was an important aspect of American political science during the 1950s and 1960s. It presented itself as a challenge to what it characterized as traditional and institutional approaches to the study of politics, and it became a contentious but increasingly ascendant position in the field. Despite its avowed commitment to a value-free stance, it continued to reflect the liberal democratic values embedded in what Bernard Crick had famously characterized as the American science of politics, in his 1958 book of that title. The term *behavioralism* was adopted by political scientists after a group of social scientists at the University of Chicago, who were devoted to developing a general scientific interdisciplinary study of human behavior, coined, in 1949, the label "behavioral sciences." The term was intended to suggest an attitude of scientific objectivity and, unlike "social science," to allay a persistent worry, revived during the Cold War, about confusion with the ideology of socialism. Although the term related to a long-standing interest, among Charles Merriam, Harold Lasswell, and others in the Chicago school of political science, in the psychological bases of politics, it did not signify any direct association with the theory of behaviorism.

Pendleton Herring, who became president of the Social Science Research Council in 1948 and who advocated the scientific study of political "behavior," had, since the 1920s, been a significant contributor to what was conceived as an empirical theory of pluralist democracy. This theory had emerged during the first quarter of the 20th century, and Herring, on the eve of World War II (in *The Politics of Democracy,* 1940), had stressed the need to ground democratic values in a scientific descriptive account of politics. This theme persisted during the early behavioral era in the work of individuals such as David Truman (*The Governmental Process,* 1951) and Robert Dahl (*Preface to Democratic Theory,* 1957; *Who Governs,* 1960), who not only advocated a more scientific mode of inquiry but also presented strong defenses of pluralist, or what Dahl referred to as "polyarchial," democracy in the United States and other Western societies.

David Easton, who came to Chicago in 1948 after finishing his degree at Harvard, became a principal theorist of the behavioral movement and the primary exponent of systems theory in political science (*A Systems Analysis of Political Life,* 1965). The first significant statement of the behavioral position was his 1951 article "Decline of Modern Political Theory," in which he criticized the "poverty" of theory in the discipline and ascribed this condition to absorption with the study of the history of past ideas and the consequent failure to take on the "task of building systematic theory about political behavior" as well as advancing relevant value theory. In 1961, Dahl claimed that the "behavioral approach" represented a "successful protest" against "conventional political science" and the work of speculative theorists, historians, legalists, and moralists and that it involved a "mood" or "outlook" that emphasized empirical modes of investigation and a focus on what

"is" as opposed to what "ought" to be. This sentiment was again apparent in Easton's 1962 statement of "The Current Meaning of 'Behavioralism,'" in which he defined it as a "science of politics modeled after the methodological assumptions of the natural sciences" (p. 17). Among the other "tenets" of the behavioral "credo" were a distinction between ethical claims and "empirical explanation" and the assumption that the achievements of "pure science" should precede the practical application of the knowledge. In the 1968 edition of the *International Encyclopedia of the Social Sciences,* Easton noted that "the methods of modern science had made deep inroads into political research, under the rubric of the study of political behavior" (p. 1295), which involved a sharp break with the past, the accumulation of large amounts of empirical data, the introduction of theoretical coherence, and a clear distinction between factual and normative claims.

A general commitment to a scientific study of politics was not an innovation in American political science, but the available philosophical accounts of science, often derived from the literature of American pragmatism, had been less than systematic, and in practice, a scientific approach often amounted to little more than data collection and the application of quantitative techniques. By midcentury, European scholars such as Rudolf Carnap and Carl Hempel, who were among the founders of the philosophy of logical positivism, had immigrated to the United States and offered a highly structured defense of the unity of science and a reconstruction of the logic and epistemology of science. This work gained hegemony in the literature of the philosophy of science, and often in secondary derivations, it served as a model for political scientists who sought both the authority of science and a guide to scientific practice. Beginning at least with Lasswell's collaboration with the philosopher Abraham Kaplan in the 1950 work *Power and Society,* it was apparent that what behavioralists meant by "science" pivoted on the positivist paradigm. For Lasswell and Kaplan, however, the purpose of pursuing a scientific study of politics was still, as for an earlier generation, to achieve the kind of epistemic credibility that would serve to bring "political theory and practical politics into closer harmony." This image of science was, however, vigorously opposed by numerous critics, including émigré scholars such as Leo Strauss and members of the Frankfurt School. The critics challenged the ideology of liberalism embedded in the behavioral program as well as what they charged was the manner in which scientism had led both to moral relativism and to a failure to confront contemporary political issues.

Only a year after his affirmation of the dominance of behavioralism, Easton, in his 1969 presidential address to the American Political Science Association, responded to the critics and announced a "new" or "postbehavioral" revolution in political science that sprung from a "deep dissatisfaction with political research and teaching, especially of the kind that is striving to convert the study of politics into a more rigorously scientific discipline" (p. 1051). Easton did not abandon the basic principles of behavioralism but instead called for a different distribution of emphasis whereby the pursuit of "the discovery of demonstrable basic truths about politics" would, in the short run, be subservient to the application of currently available knowledge to the "problems of the day." This new "policy turn" had begun, however, to fade by the end of the 1970s, and the underlying methodological agenda of behavioralism continued significantly to inform the future evolution of the field.

John G. Gunnell

See also Behaviorism, Philosophical Conception of; Behaviorism in Psychological Explanation; Empiricism; Frankfurt School and Critical Social Theory; Logical Positivism/Logical Empiricism; Straussian Critique of Social Science; Systems Theory

Further Readings

Easton, D. (1962). Introduction: The current meaning of "behavioralism." In J. C. Charlesworth (Ed.), *Political science* (pp. 8–25). Philadelphia, PA: American Academy of Political and Social Science.

Easton, D. (1968). Political science method and theory. In D. Sills (Ed.), *International encyclopedia of the social sciences* (Vol. 12, pp. 1295–1297). New York, NY: Macmillan.

Easton, D. (1969). The new revolution in political science. *American Political Science Review, 63,* 1051–1061.

Farr, J. (2003). Political science. In T. M. Porter & D. Ross (Eds.), *The Cambridge history of science: Vol. 7. The modern social sciences* (pp. 306–328). Cambridge, England: Cambridge University Press.

Farr, J., & Seidelman, R. (Eds.). (1993). *Discipline and history: Political science in the United States.* Ann Arbor: University of Michigan Press.

Gunnell, J. G. (1993). *The descent of political theory: The genealogy of an American vocation.* Chicago, IL: University of Chicago Press.
Gunnell, J. G. (2004). *Imagining the American polity: Political science and the discourse of democracy.* University Park: Pennsylvania State University Press.
Sigelman, L. (Ed.). (2006). The evolution of political science [Centennial issue]. *American Political Science Review.*
Somit, A., & Tanenhaus, J. (1967). *The development of American political science: From burgess to behavioralism.* Boston, MA: Allyn & Bacon.

Behaviorism, Philosophical Conception of

According to behaviorism, everything there is to know or say about people with regard to their mental states can be known or said in terms of their observable behaviors (including verbal behaviors), and furthermore, all there *is* to mental states themselves are certain patterns of behavior or dispositions to behave. There are three core claims of behaviorism: (1) an epistemological claim (mental states are knowable only via behavior), (2) a semantic claim (mental-state words like *belief* have meanings definable in terms of behavior), and (3) a metaphysical claim (either there are *no* mental states, just behaviors and dispositions to behave, or there *are* mental states but they are identical to behaviors or dispositions to behave). Behaviorism as a movement in philosophy overlapped partially with a movement of the same name in psychology. This entry discusses philosophical behaviorism by discussing its history and some of the major arguments, both pro and con.

History of Philosophical Behaviorism

Philosophical behaviorism emerged from logical positivism and ordinary-language philosophy. Positivists believed in verificationism, according to which the meaning of a term is given by specifying the observable conditions that would verify its application. Ordinary-language philosophers were suspicious of philosophical theses, such as dualism, that were not stated in the terms of ordinary language. In the philosophy of mind, especially in the 20th century, many philosophers saw behaviorism as a viable early contender for a materialistic solution to the mind–body problem. Few contemporary philosophers of mind attracted to mind–body materialism subscribe to behaviorism, preferring instead one of its successors, such as functionalism and the mind–brain identity theory. While controversy surrounds applying the label of "behaviorist" to the philosophers who are often cited as being behaviorists, such citations occur frequently enough to merit mentioning as key behaviorists the philosophers Gilbert Ryle, Ludwig Wittgenstein, W. V. O. Quine, and Daniel Dennett.

Motivating Behaviorism

The private-language argument from Ludwig Wittgenstein's *Philosophical Investigations* has behavioristic conclusions. Wittgenstein attacked the alleged privacy of mental states. The conclusion of Wittgenstein's argument is that it is impossible for there to be a language that referred only to private things: a language about sensations that could only be understood by a single person. Suppose you devise a language in which there is a sign, "S," that you intend to stand for a particular sensation. According to Wittgenstein, no one, not you or anyone else, can distinguish between a correct usage and a mere *seemingly correct* usage of "S." However, where one cannot grasp a distinction between correct and incorrect uses, there is no place for a notion of correctness at all. Thus, "S," as well as the rest of the signs in this so-called private language, is meaningless.

Another line of thought with behavioristic conclusions is based on the verificationism central to logical positivism, according to which terms are defined by the evidence for their correct application. Given that the main evidence we have for correctly applying terms like *belief* and *desire* is behaviorial, verificationism entails that the meaning of such terms is definable in terms of behavior.

Against Behaviorism

One objection to behaviorism hinges on qualia, the subjective aspects of our conscious mental states, especially sensory states like seeing a red rose. If behaviorism is true, then it ought to be inconceivable for two beings to share all their behavioral dispositions but to differ in what qualia accompany their sensory interactions with roses (e.g., the way roses look to one differs from the way they look to the other). However, since such a situation is conceivable, behaviorism is false.

A second objection to behaviorism is based on the claim that it is part of our concepts of mental states that they can serve as causal explanations of behavior. For instance, one's opening of a refrigerator is explained causally by appeal to one's beliefs and desires (e.g., a desire to drink beer and a belief that there is beer in the refrigerator). However, by defining "belief" and "desire" by reference to such behaviors, behaviorism renders such explanations unacceptably circular.

A third objection to behaviorism is based on the claim that mental states cannot be individually connected with behaviors but can only be connected to behaviors in concert with other mental states. Whether a person's desire to avoid tiger attacks will result in her running away from a tiger instead of toward it (or not running at all) depends on her beliefs about where the nearest tiger is and whether it is more likely to attack a stationary person or a running one. The problem is that the project of saying which behavior a mental state is connected to is so complicated as to be totally intractable. Worse, the project of behavioral definition is thereby shown to be circular. Each mental state can only be connected to behavior by reference to other mental states, including the mental state that we started with, and thus we are led in a circle.

A fourth objection is the *perfect-actor* objection. For any set of behaviors one might engage in while in a mental state, a perfect actor can perform those very same behaviors without being in that mental state. Well-trained actors know techniques to help them appear sad even when they aren't actually sad. Thus, for any given mental state, no set of behavioral dispositions is essentially associated with it.

Pete Mandik

See also Behaviorism in Psychological Explanation; Logical Positivism/Logical Empiricism; Mind–Body Relation; Philosophical Psychology, History of; Verificationism

Further Readings

Dennett, D. (1978). Why the law of effect will not go away. In D. Dennett (Ed.), *Brainstorms: Philosophical essays on mind and psychology* (pp. 71–89). Cambridge: MIT Press.

Quine, W. V. O. (1960). *Word and object.* Cambridge: MIT Press.

Ryle, G. (1949). *The concept of mind.* Chicago, IL: University of Chicago Press.

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). New York, NY: Macmillan.

Behaviorism in Psychological Explanation

In 1890, America's first great psychologist, William James, famously proclaimed psychology to be the science of mental life, whose main aim is to determine the correlation of various kinds of thoughts and feelings with brain activity. Despite the prominence of James's proclamation, a competing school of psychology—*behaviorism*—emerged less than a quarter of a century later, which represented a seismic shift away from psychology as a science of mental life and toward a science of overt action.

From Consciousness to Behavior

No matter how compelling conscious mental experience was to James, rival theorists such as Ivan P. Pavlov, H. S. Jennings, and John B. Watson deemed such personal experience to be as useless as data for a scientific psychology as it would be for physics or chemistry. Why? It is not because behaviorists flatly deny the very existence of consciousness; rather, it is because they believe that private experience cannot be the proper subject of public science. In an oft-quoted expression, the mind is considered by radical behaviorists, like B. F. Skinner, as a "black box" in the sense that an alleged "internal" or "private" psychological entity, such as a conscious mind or its neural workings, could not be studied by science proper.

Can I see it? Can I measure it? Can I repeat my results? These three questions were to become a veritable mantra to Pavlov and his students in the aftermath of their initial, futile attempts to provide a "psychical" account of what they later termed conditioned reflexes. A natural science of behavior could not indulge the naive interpretation of animal behavior in terms of conscious human experience; what was necessary to explain both human and animal behavior was an objective account that sought the physiological mechanisms of what Pavlov called "higher nervous activity."

From Humans to Animals

It is surely no accident for the development of behaviorism that Pavlov, Jennings, and Watson had all assiduously studied the behavior of animals. With introspection thereby taken off the methodological table, overt behavior would have to be the focus of psychology as a natural science. So for behaviorists, the objective processes in behavior became of greatest interest in their own right. Discovering general laws of behavior—in both humans and animals—with the methods of natural science became the goal of behavioristic psychology.

Behaviorists saw this extrapolation of psychological science to animals to be especially important. It not only broadened the scope of scientific inquiry, but it also suggested that evolutionary principles of behavior might be as effectively applied to humans as to animals. This evolutionary extension was also controversial because it undermined the orthodoxy of the Cartesian belief that humans differed dramatically from all other animals. From a behavioristic perspective, it would no longer do to explain human behavior in mental terms and to explain animal behavior in physical terms. Only by objectively comparing the overt actions of both humans and animals can we effectively ascertain whether there is a continuity or a gulf between them. Charles Darwin unlatched the door to the possibility of "continuity" between humans and animals. Behaviorists flung it wide open and set their sights on collecting unimpeachable empirical evidence to rule on this provocative possibility.

From Theory to Practice

Behaviorists have been motivated not only by theoretical concerns but by practical problems as well. How might abnormal and ineffective human behaviors be best comprehended and remedied? Behaviorists have responded by carefully determining the origins of these behaviors and by devising ways of adaptively modifying the behaviors. Initially, Pavlov and Watson and, later, Skinner applied the principles of behavioral analysis to improving the human condition.

For example, behavior therapy seeks to alleviate behavioral dysfunctions by modifying the controlling conditions. Effective treatments for anxiety, enuresis, self-injurious behavior, stuttering, bulimia, depression, insomnia, and many other disorders have arisen from the basic scientific research of behaviorists. Also, applied behavioral analysis has proven to be successful in preventing the transmission of AIDS, promoting industrial safety and seat belt use, conserving natural resources, reducing littering, improving educational practices, and encouraging health and exercise.

Criticisms and Rejoinders

Despite these many important theoretical and practical achievements, many psychologists and philosophers now insist, as they have for several decades, that behaviorism is dead. Yet, much like Mark Twain's premature obituary, "the reports of behaviorism's death are greatly exaggerated."

What are these alleged failings, and why might they not be true?

Critics often contend that behaviorism offers an impoverished view of the individual and society, with no place for perceptual, cognitive, emotional, and spiritual mechanisms. In particular, how can any science of mind ignore consciousness? Behaviorists counter that objective science must deal with the observable. Such a focus may set aside spiritual and metaphysical matters, but the study of perception, cognition, and emotion remains very much on the table when viewed as behavior rather than as personal experience. Indeed, notable progress has been made with advanced behavioral methods in understanding the mechanisms of perception, cognition, and emotion in both humans and animals.

Another complaint is that many of our distinctively human qualities, like language, have an innate foundation; because behaviorism focuses on acquired actions, it can say little about many of those mental attributes that so decidedly distinguish us from animals. This claim too is incorrect. Even in its infancy, behaviorism appreciated the intricate interplay between ontogeny and phylogeny in the behavior of both humans and animals. Indeed, the very notion of innate behavior is undergoing a systematic reanalysis in light of recent research in developmental psychobiology.

A further rebuke is that behaviorism, as a deterministic approach to psychology, makes people into automatons lacking free will. Of course, humans and animals are subject to many illusions: Free will may be the most compelling. More to the point, behaviorists not only accept that life offers organisms

innumerable choices in behaviors and reinforcers, but they believe that there are lawful regularities in those choices that can be scientifically discovered. We surely are a long way from being able to say at any given time which of several choices an organism will make, but the hope is that a science of behavior is learning to predict such choices with increasing accuracy. What scientific strategy would critics propose instead?

In summary, behaviorists contend, "We are what we do." Understanding why we do what we do is the scientific quest.

Edward A. Wasserman

See also Behavioralism in Political Science; Behaviorism, Philosophical Conception of; Empiricism; Logical Positivism/Logical Empiricism; Mind–Body Relation

Further Readings

Bekoff, M., Allen, C., & Burghardt, G. M. (Eds.). (2002). *The cognitive animal.* Cambridge: MIT Press.

Blumberg, M. S. (2005). *Basic instinct: The genesis of behavior.* New York, NY: Thunder's Mouth.

Blumberg, M. S., & Wasserman, E. A. (1995). Animal mind and the argument from design. *American Psychologist, 50,* 133–144.

Griffin, D. R. (2001). *Animal minds.* Chicago, IL: University of Chicago Press.

Jennings, H. S. (1976). *Behavior of the lower organisms.* Bloomington: Indiana University Press. (Original work published 1906)

Mackenzie, B. D. (1977). *Behaviourism and the limits of scientific method.* Atlantic Highlands, NJ: Humanities Press.

Pavlov, I. P. (1928). *Lectures on conditioned reflexes.* New York, NY: International Press.

Roediger, R. (2004). What happened to behaviorism? *Association for Psychological Science: Observer, 17*(5), 40–42. Retrieved from http://www.psychologicalscience.org/observer/getArticle.cfm?id=1540

Skinner, B. F. (1976). *About behaviorism.* New York, NY: Vintage.

Sober, E. (1983). Mentalism and behaviorism in comparative psychology. In D. W. Rajecki (Ed.), *Comparing behavior: Studying man studying animals* (pp. 113–142). Hillsdale, NJ: Lawrence Erlbaum.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review, 20,* 158–177.

Wegner, D. M. (2003). The mind's best trick: How we experience conscious will. *Trends in Cognitive Science, 7,* 65–69.

Zuriff, G. E. (1985). *Behaviorism: A conceptual reconstruction.* New York, NY: Columbia University Press.

Being-in-the-World

"Being-in-the-world" (*In-der-Welt-sein*) is the term of art coined by Martin Heidegger in his 1927 magnum opus, *Being and Time,* to describe the kind of existence specific to us, namely human beings, whom he calls *Dasein* (an otherwise ordinary German word meaning "being" or "existence").

The philosophical significance of Heidegger's use of these terms rests on two original and controversial claims. The first is what he calls the "ontological difference" between "entities" (*das Seiende*) and "being" (*das Sein*). Entities are whatever can be said *to be.* By contrast, that *in virtue of which* an entity is an entity, and that on the basis of or in terms of which we understand an entity to be, is its being. Heidegger's reason for drawing this distinction is to insist that being is not itself an entity and so cannot be investigated in the way the natural and human sciences investigate their objects. The question of being belongs exclusively to philosophy.

The second philosophical innovation behind Heidegger's introduction of the term *being-in-the-world* is his insistence that there are not just one but several ways or modes of being. That is, not only are there many things, and indeed many different kinds of things, but also several distinct and irreducible kinds of existence: What it means *to be* is not the same for everything that can be said to be. So, for example, what it means to be an "occurrent" (*vorhanden*) object (a stone, a tree, a molecule, a planet) is simply to be present at some particular point in time. By contrast, what it means to be an "available" (*zuhanden*) tool or piece of equipment (a shoe, a chair, a train, a coin) is to be useful in some practical context. Something's being *available* for use, Heidegger argues, is not just a special case of its being *occurrent* and then also having some special set of use properties added on. Rather, *being* available and *being* occurrent are two fundamentally different ways of being.

Dasein is neither occurrent nor available. Human beings are neither objects nor functional systems with human features in addition. More generally, a

human being is not a *what* but a *who*. For a human being to be is for it to inhabit a social and historical world in which it has intelligent skills to attend to things, orient itself, move, and undertake actions. The human being is a *being-in-the-world* (for which Heidegger also uses the German word *Existenz*). The central thesis of *Being and Time* is that each kind of being is intelligible in terms of its own kind of time—the instantaneous *now* for occurrent entities (as measured by clocks, for example) and the situational, pragmatic *now* for available things (time to eat, time to sleep). Human existence is temporal in a way fundamentally different from either of these two, for we are not confined to a discrete *now* or even to a flowing sequence of past, present, and future *nows*. Instead, our existence consists in our always having a future, understood as a horizon of possibilities into which we are constantly pressing or "projecting," thinkingly or unthinkingly, and a past, understood as the *already* of the given situation that we are "thrown" into, willingly or unwillingly.

Dasein's future-constituting possibilities are disclosed to it in what Heidegger calls "understanding" (*Verstehen*), by which he means competence or *know-how,* which includes not just intellectual capacities but also bodily skills, such as the ability to hammer a nail or ride a bicycle. Cognitive *knowing-that* is derivative of and dependent on practical and existential understanding. In addition to its propositional beliefs and judgments, we always have a pretheoretical, noncognitive understanding of our world, of our projects and commitments, and of ourselves and our being. *Dasein's* existential understanding embeds it in a practical world structured by a teleological significance constituted by means, goals, and ultimate meaning-giving ends, the point or "for-the-sake-of-which" we do what we do. The "world" of being-in-the-world is thus not a mere collection of objects but a familiar place where one can live or "dwell" (*wohnen*).

In addition to our projective understanding, we also have moods and emotions. That is, we are always affectively attuned to ourselves and to our world as an already constituted situation into which we are thrown. *Dasein* is exposed to its "thrownness" not by understanding but by mood or affect, which Heidegger calls "disposedness" (*Befindlichkeit*). Crudely speaking, whereas projection is active, thrownness is passive, and neither dimension of being-in-the-world is reducible to the other. Certain fundamental moods—above all, "anxiety" (*Angst*), but also joy and boredom—reveal the world to *Dasein* as a whole, including itself and its own being.

Human existence is also essentially social: Being-in-the-world, Heidegger maintains, is always also "being-with" (*Mitsein*). The philosophical problem of solipsism and other minds is a perverse effect that is misconceiving of *Dasein* as a "worldless" epistemological subject. We do not encounter others as alien subject–object anomalies in our own solitary world; instead, our most basic self-understanding is already an understanding of ourselves as among others and as defined by social norms. The world, Heidegger says, is always deeply articulated by "the one" or "anyone" (*das Man*) of normal, appropriate thought and behavior: That is, for the most part, we do what "one" does.

Being-in-the-world is also essentially "being toward death" (*Sein zum Tode*). Death in the existential sense is not an accidental occurrence that takes place at the end of our lives but a collapse of meaning and possibility that Heidegger believes is essential to the finitude of our existence throughout the duration of our lives.

Heidegger argues that the meaning of being-in-the-world is "care" (*Sorge*). For human beings, to be just is to care about something, for something to matter to us, that is, to be affected in mood and to project into the world as a domain of significance. Taking his inspiration from Heidegger, Hubert Dreyfus has criticized research on artificial intelligence by arguing that computers cannot exhibit anything like human understanding as long as they lack bodily skills and affective care.

Dasein can exist either "authentically" or "inauthentically," that is, either *owning up* to or fleeing from and effectively *disowning* its situation and itself in alienation and social conformism. "Authenticity" (*Eigentlichkeit*) involves what Heidegger calls "running forth into death," which means not suicide but embracing one's finitude in "resoluteness" (*Entschlossenheit*), or wholehearted commitment to what one cares about.

Taylor Carman

See also Embodied Cognition; Existential Phenomenology and the Social Sciences; Hermeneutics, Phenomenology, and Meaning; Intersubjectivity; Life-World; Phenomenological Schools of Psychology; Scheler's Social Person; Tacit Knowledge

Further Readings

Blattner, W. (2006). *Heidegger's Being and Time: A reader's guide*. New York, NY: Continuum.

Carman, T. (2003). *Heidegger's analytic: Interpretation, discourse, and authenticity in Being and Time*. Cambridge, England: Cambridge University Press.

Dreyfus, H. L. (1991). *Being-in-the-world: A commentary on Heidegger's Being and Time, division I*. Cambridge: MIT Press.

Heidegger, M. (2008). *Being and time* (J. Macquarrie & E. Robinson, Trans.). New York, NY: Harper & Row. (Original work published 1962)

Wrathall, M. (2006). *How to read Heidegger*. New York, NY: W. W. Norton.

Biology and the Social Sciences

This entry charts the historical and thematic development of the conceptual relationship between biology and the social sciences.

Before World War II

Social science in the contemporary institutional sense began in the last decades of the 19th century. Of course, numerous philosophers and others before that time had thought systematically about the structures of human social interactions. Many, including Plato and Aristotle, appealed to biological models and metaphors in doing so. Before the invention and spread of industrial manufacturing, organisms with their specialized but integrated parts were the only kinds of objects that seemed to parallel human societies in functional complexity. Thus, before the industrial revolution, organic conceptions of society that depicted certain common forms, such as cities, as natural—rather than conventional or engineered—were dominant. The most influential model of this kind was that of Bernard Mandeville, whose *Fable of the Bees* (1705 and 1714) is generally regarded as an early literary expression of the idea of the invisible hand that preserves social order and economic productivity in ways that no individual intends or deliberately tries to produce. It is only when the world began to be economically dominated by production using industrial machinery that increasing numbers of thinkers started depicting unsocialized individuals as natural and social institutions as artificial constructions.

Social reflection also had an important reverse influence on biological thinking. The most important single stimulus to Charles Darwin's recognition of evolution by natural selection, the great organizing principle of all modern biology, was the work of the demographic theorist Thomas Malthus. In 1826, Malthus argued that human population was limited by the finite food-carrying capacity of available land and that lives would become miserable as this limit was approached. Darwin noticed that this reasoning implied competition and that less successful competitors would tend to be extinguished under the resulting selection pressure, taking with them the features that made them less competitively successful. If resources remained finite, such competitive pressure would always operate even if each generation were more efficient in resource utilization than the preceding one. Thus, Darwin concluded, both the natural and the social ecologies would tend to continuously evolve.

This Darwinian conception then fed back into social thinking. Social Darwinists, such as A. R. Wallace and Herbert Spencer, and normative organicists, such as Ernst Haeckel, maintained that winners of competitions to dominate society by commercial or political advantage thereby tended to improve society, just as the so-called selection of the fittest drove improvements in the adaptedness of successive plant and animal species. In this view, they misunderstood the more careful Darwin, who saw that natural selection was not necessarily a force for qualitative progress if environments continuously change, as they do. Darwin's understanding was too subtle for most nonbiologists, and so his science was widely invoked in the early 20th century by social reformers, both well-intentioned and otherwise, who promoted eugenics policies aimed at discouraging supposedly "inferior" kinds of people from reproducing.

Antibiology Backlash

The most extreme version of eugenics was practiced during World War II, when the Nazi leadership of Germany set out to exterminate whole populations of people—mainly Jews—whom they believed to be biologically distinctive from themselves and, on that basis, to also be morally inferior. The Japanese military government adopted a similar view and policy

with respect to the Chinese and others in Asia whom they briefly conquered. These terrible events, based partly on rampant confusion about human biology, led to a backlash in the years following World War II against all appeal to biological models or influences in conceptualizing the structure of society or the forces that drive social change.

This initial negative basis for a turn toward promoting the complete independence of the social from the biological was complemented by a positive idea from the ideological left. Some biologists in the Soviet Union had promoted the thesis—based on no sound evidence—that changes in the capacities of an organism could be inherited by its offspring. This encouraged optimism about the pace at which enduring revolutionary changes in human behavioral tendencies could be effected. Although this outright denial of basic Darwinism died with the Soviet leader Joseph Stalin in 1953, many promoters of radical social restructuring throughout the world continued to urge the autonomy of the social from the biological. This rested partly on misconceptions about nonhuman animals that were widespread among social scientists until comparatively recently: that such animals are genetically programmed to exclusively promote their individual welfare, that their social responses are stereotyped and shallow, and that there is little personality or other psychological variation among the members of any given nonhuman species. Thus, all the key preconditions for dynamic cultural and political change were widely considered to be missing in the biological domain.

Sociobiology and Evolutionary Psychology

The above reasoning takes optimism about humanity's capacity for change as an assumed premise. In the mid-1970s, a number of thinkers, including most famously the entomologist E. O. Wilson, called this optimism into question by drawing attention to numerous ways in which human habits seem to be anchored to, and to be explicable only by reference to, fitness maximization of social ground-dwelling apes. For example, humans seem intractably inclined to tribalism, the tendency to strongly favor small groups of genetic relatives and associated friendship networks, even in circumstances where more cosmopolitan attitudes would be more generally advantageous. During the decades since Wilson's promotion of this idea, it has been greatly enlarged upon through laboratory experiments conducted by behavioral economists. Combining testing methods from psychology with the use of monetary incentives special to economics, these researchers have examined the extent to which people identify self-interested rational solutions to strategic and other optimization challenges when forced to use their innate cognitive resources in relatively isolated and novel settings. The result has been a parade of respects in which people typically fall back on standard routines that are not ideally targeted to the unfamiliar tasks but that would make sense as evolved built-in biases given the ranges of problems that might have confronted our earlier hominid ancestors on the African savannah of the Pleistocene era.

Sociobiology was received as intensely controversial, not only scientifically but also politically and morally. This was partly because it seemed to imply limits on the transformability of society, but also more specifically because it could be taken as suggesting that some distinctions between the social roles of men and women are natural and permanent rather than culturally constructed and subject to intentional reform. These controversies continue, often under the label of the "nature/nurture debate." However, complete rejection of the idea that at least some statistically prevalent human behavioral dispositions and forms of social organization have their roots in the circumstances of hominid biological evolution is now seldom heard or taken seriously by most social scientists. At the same time, the label *sociobiology* has itself faded from usage. Since the 1990s, Wilson's methodological successors have referred to themselves as *evolutionary psychologists*. The main difference they emphasize in this rebranding is that they are clearer in distinguishing past adaptiveness of evolutionary biases from present normative functionality. (Arguably, Wilson was always equally clear about this.) More substantively, evolutionary psychologists develop their hypotheses on the basis of models of mental architecture from cognitive science. The human mind, they suggest, is not an all-purpose, rational problem-solving engine but a collection of special-purpose computers ("modules") that are activated to guide behavior through particular tasks cued by environmental triggers—especially social triggers. Failures of optimization are often diagnosed on the basis of imperfect integration among, or outright conflict between, these modules.

This picture of the mind is far from a consensus view. Some philosophers and social scientists argue that much of the supposedly discrete inner structure inferred from behavioral evidence by evolutionary psychologists involves projection in the wrong direction and, in fact, represents the recurrent sculpting of response patterns by strongly entrenched cultural articulation, to which individuals adapt during socialization. In effect, both the "nature" and the "nurture" sides of the debate have been driven by their dialectic to increasingly emphasize complexity: complexity of mental organization on the "nature" side and complexity of social structure on the "nurture" side.

Evolutionary Social Science

Despite their differences, both evolutionary psychologists and their critics tend to agree that evolutionary processes are critical to successful explanation of human social structures and social history. Evolutionary psychologists emphasize pressures of natural selection on individual genomes; evolutionary developmental theorists instead focus on the coevolution of genetic and cultural channels, operating simultaneously on individual organisms and on social structures. This has given rise to the development of "thick" historical narratives as a form of social explanation, which differ from similarly rich and historicist classic social science (e.g., Adam Smith, Alexis de Toqueville, Karl Marx) in being crucially informed by Darwinian modeling principles and, in many instances, by simulations in computer-generated dynamic virtual environments. In such so-called agent-based modeling, the researcher specifies behavioral rules that govern different types of simple individual agents, designs virtual environments that are idealizations of historical conditions of interest, and then programs rules that determine the relative success of agent types in reproducing themselves in subsequent generations, depending on the outcomes of their interactions with the environmental parameters and with one another in the current generation. In some such simulations, the evolving dynamics can also feed back to alter the background environment. (Biologists refer to this as *niche construction.*) Social scientists who build such models typically use evolutionary game theory as the mathematics for describing the relative stability of different combinations of agents, environments, and selection rules. Relatively stable, such combinations are at equilibrium. Causal dynamics of greatest interest are recurrent patterns of changes in variables that tip a simulated world out of one equilibrium and send it toward another one.

Evolutionary models of this kind, whether simulated or merely represented by equations on paper, are used both as exercises in the separate social sciences and as products of unified social science. For example, if the changing state variables in a model are relative exchange values and the principles of selection are relative asset balances measured using these values, then the model is an instance of evolutionary economics. Another model might vary the extent to which individuals form bonds with one another through spending resources on symbolic displays; this would be a case of evolutionary anthropology. Such models can be combined, creating unified social science. A major criticism of this approach is that as models become more complex and sensitive to specific simulation design parameters, they surrender the key scientific goal of generalizability.

Social Neuroscience and Neuroeconomics

The most important current arena of interaction between biological and social-scientific modeling has arisen thanks to new noninvasive technologies for probing and creating images of neural activity, such as functional magnetic resonance imaging. This has lately produced a wave of interest in the neuroprocessing basis of social affiliation (social neuroscience) and in the mechanisms by which brains compute and comparatively evaluate alternative rewards (neuroeconomics). Critics of these approaches caution against reductionism; emergent social structures might often restrict behavioral patterns to particular channels that then don't need to be explicitly represented in anyone's brain. It seems likely that this issue will dominate debates about the relationship between biological and social structures and processes over the years immediately ahead.

Don Ross

See also Agent-Based Modeling and Simulation in the Social Sciences; Cooperation, Cultural Evolution of; Cultural Evolution; Evolutionary Game Theory and Sociality; Evolutionary Political Science; Evolutionary Psychology; Human Cultural Niche Construction and

the Social Sciences; Neuroeconomics; Primatology and Social Science Applications; Social Neuroscience; Sociobiology

Further Readings

Barkow, J., Cosmides, L., & Tooby, J. (Eds.). (1992). *The adapted mind*. Oxford, England: Oxford University Press.

Boyd, R., & Richerson, P. (1988). *Culture and the evolutionary process*. Chicago, IL: University of Chicago Press.

Cacioppo, J. (Ed.). (2002). *Foundations in social neuroscience*. Cambridge: MIT Press.

Epstein, J., & Axtell, R. (1996). *Growing artificial societies*. Cambridge: MIT Press.

Glimcher, P., Camerer, C., Fehr, E., & Poldrack, R. (Eds.). (2009). *Neuroeconomics: Decision making and the brain*. Amsterdam, Netherlands: Elsevier.

Laland, K., & Brown, G. (2002). *Sense and nonsense: Evolutionary perspectives on human behavior*. Oxford, England: Oxford University Press.

Sterelny, K. (2003). *Thought in a hostile world*. Oxford, England: Blackwell.

Wilson, E. O. (1975). *Sociobiology: The new synthesis*. Cambridge, MA: Harvard University Press.

C

Capabilities

This entry introduces what is known as the capability approach and goes on to review various senses and uses of the notion of capability in philosophy, the social sciences, and policy.

The term *capability* is typically used to denote a potential, that is, the ability to achieve something that is possessed by a given entity. In philosophy and in the social sciences, the term is often used in connection with the "capability approach," proposed by Martha Nussbaum and Amartya Sen. The capability approach was first proposed in 1979 by Amartya Sen in a lecture titled "Equality of What?"—in which Sen tries to go beyond John Rawls's theory of justice and its use of "primary goods" as the relevant space in which to assess inequality.

Sen argues that by focusing on "primary goods," Rawls fails to take into account the differences between human beings. Different human beings will have different abilities to convert primary goods into well-being. Thus, Sen and Nussbaum argue that equality and well-being should be assessed in terms of the capabilities that human beings actually possess, which depend not only on the goods and resources they possess but also on the conversion factors that enable human beings to convert goods and resources into well-being.

Sen notes that there are other approaches that take into account the differences between human beings, such as utilitarianism (which underpins mainstream economics). However, Sen criticizes the use of utility as the space in which to assess equality and well-being. Sen argues that utility is a subjective measure, which depends on subjective preferences and may not reflect human well-being. For example, if our preferences become adapted to a given situation, our utility level may increase, but that subjective phenomenon does not mean that there is an increase in well-being. Thus, human well-being should be assessed in terms of the human functionings we have reason to value, where a human functioning is what a given human being is or does. The notion of functioning has Aristotelian roots; it has been developed especially by Nussbaum but has also been acknowledged by Sen.

Furthermore, the capability approach provides a multidimensional perspective on human well-being, since it focuses on various human functionings. However, Sen argues that equality should be assessed taking into account not only achieved functionings but also the *potential* to achieve. Thus, equality should be assessed in the space of the *potential* functionings that we have reason to value, which Sen and Nussbaum designate as valuable *capabilities,* that is, what a human being can be or do and have reason to value.

The capability approach has been adopted in developing concrete measures of well-being, most notably by the United Nations Human Development Programme, which publishes every year a Human Development Report, where the multidimensional approach of the capability approach is employed, going beyond the traditional measures that rely on the gross domestic product. The capability approach

also became influential in feminist studies and within political philosophy.

In business and management studies, the term *capability* is more often identified with the capabilities theory of the firm. This approach goes back to Edith Penrose and is connected to the resource-based view of the firm, where the term *capability* refers to the ability of a firm to use resources. Like in Sen's and Nussbaums's capability approach, the aim is to go from goods and resources to capabilities. However, in the theory of the firm, capabilities are seen as relevant to the competitive advantage of the firm rather than to assessing equality among human beings.

Capabilities also depend on the firms' routines, a concept developed within evolutionary perspectives, such as that of Robert Nelson and Sidney Winter, where the capabilities of the organization (organizational capabilities) depend on the routines of the organization (organizational routines).

In the theory of the firm, the term *competences* is also often used to denote a concept similar to that of "capability." Although much use of structural terms is made in the literature on the theory of the firm, there is little systematic account given of the different kinds of structural properties at play. One relatively unknown but potentially helpful account is given by Clive Lawson, who proposes a useful distinction between *capacity, capability,* and *competence*. A *capacity* is a more passive power of a structure to achieve something, a *capability* refers to an ability that has not yet been acquired, and a *competence* refers to an ability that has been acquired.

But across the various literatures on management, philosophy, and the social sciences, the term *capability* always denotes a potential that a given entity (such as a human being or an organization) possesses.

Nuno Martins

See also Agency; Decision Theory; Homo Economicus; Reflective Equilibrium; Social Choice Theory

Further Readings

Lawson, C. (2004). Competence theories. In S. Fleetwood & S. Ackroyd (Eds.), *Critical realist applications in organisation and management studies* (pp. 234–252). London, England: Routledge.

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Cambridge, MA: Belknap Press.

Nussbaum, M. C. (1992). Human functioning and social justice: In defense of Aristotelian essentialism. *Political Theory, 20,* 202–246.

Nussbaum, M. C. (2000). *Women and human development: The capabilities approach*. Cambridge, England: Cambridge University Press.

Nussbaum, M. C., & Sen, A. K. (Eds.). (1993). *The quality of life*. Oxford, England: Oxford University Press.

Penrose, E. T. (1959). *The theory of the growth of the firm*. Oxford, England: Basil Blackwell.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Belknap Press.

Sen, A. K. (1982). *Choice, welfare, and measurement*. Cambridge: MIT Press.

Sen, A. K. (1999). *Development as freedom*. New York, NY: Alfred A. Knopf.

Sen, A. K. (2009). *The idea of justice*. Cambridge, MA: Harvard University Press.

Capitalism

Capitalism is an idealized economic system consisting of legal protections of decentralized, cooperative social production; private ownership of resources; and free and open competitive markets for exchange of goods, labor, services, and material and financial capital. Capitalism can be contrasted with socialism, which promotes collective ownership of productive resources and labor, and with traditional, preindustrial economies, which determine production and distribution of goods and labor by traditional social roles. Capitalism dominates globally, promoting worldwide trade and labor mobility, but markets exist only in and through legal systems that are bounded by nation-states. These legal systems, along with the social norms that support and constrain each society's economy, create different approximations to the ideal of capitalism in different nations. While Karl Marx coined the term *capitalism* as part of a broader economic theory of production and distribution, neoclassical economic theory generally assumes the existence of capitalist markets, but heterodox economics and sociological and normative political theories provide a critical distance to critique capitalism as a system.

Private Property and Free Markets

Property rights are defined as the legal rights to possess, consume, use, trade, or destroy goods. Capitalism exists only where those legal rights allow private individuals to use goods (including material, intellectual, and financial goods), and to hire laborers to produce other goods, for trade. In a capitalist system, consumers

ultimately determine what is produced through their individual consumption decisions, and producers must compete against each other to appeal to consumer desires. Competition among producers leads to business failures and the creation of new businesses vying to create products that consumers will buy. This process, which the economist and political scientist Joseph Schumpeter called "creative destruction," leads not only to innovation in consumer goods but also to financial ruin and unemployment, at least temporarily, for producers and workers. It also leads to unequal ownership of goods among individuals and to periodic social upheavals. Socialism prohibits private ownership of large-scale production, which requires government intervention in productive and market activities. When property ownership, production, or trade is very highly taxed, a market-based system can be a hybrid between capitalism and socialism, known as *market socialism.* Variants of (pseudo) capitalism include *state capitalism,* where the state supports certain industries against competition, and *oligarchic capitalism,* where the means of production are owned exclusively by a small sector of society, determined by forces outside the market. Without a market system to allow individual choice to determine what will be produced and who will consume it, such decisions must be made and enforced collectively. Although it was a commonly held view among economists until the mid 20th century that such decisions could be effectively coordinated, the collapse of communist economies has cast serious doubt on that proposition.

Global Trade

Capitalism and the concomitant growth of industry and transportation have accelerated the global spread of culture, wealth, and persons through trade. Although property rights and markets are governed by the legal systems of the countries within which trade takes place, when trade crosses national boundaries these transactions are governed by bilateral and multilateral agreements between and among the participating nations. Given the different conceptions of just social distributions and different social norms and beliefs about legitimate rules for consumer and labor protection, disagreements about property rights and market restrictions inevitably arise between nations. The basic tenet of welfare economics holds that trade increases overall welfare and that as long as it is unforced, it cannot decrease any individual trader's welfare. However, trade agreements between nations necessarily benefit some nations and some individuals more than others, creating internal and external friction. In the post–World War II era, transnational institutions were developed to reduce tariffs and trade wars and foster international economic growth and development. These institutions include the World Trade Organization, which adjudicates trade disputes between nations; the International Monetary Fund, which is the global lending organization of last resort; the World Bank, which aims at promoting economic development and poverty reduction in the developing world; and the International Labor Organization, which promotes the rights of laborers. Human rights questions and disputes that arise in the context of global capitalism, such as in the matter of human trafficking, are scrutinized by the United Nations.

Controversies of Global Capitalism

Normative and social scientific study of capitalism raises many urgent questions about how capitalism relates to economic development, human rights, the growth of inequality, and the destruction of the natural environment and traditional cultures. Capitalism has been the dominant economic system during a time of unparalleled growth in material well-being for much of the world. Life expectancy has gone from around 30 in 1800 to upward of 65 years in nearly every nation, and is higher than 80 years in most of the developed world. Child mortality and women's fertility rates have likewise declined precipitously. Claims about moral progress are less easily quantified, but the past century of global capitalist domination has seen the end of legal slavery and the enactment of formal, legal rights for women in much of the world. Capitalism by its nature promotes social mobility, in that traditional forms of status do not determine the social role of individuals from birth. However, despite these positive gains, massive absolute poverty and inequalities in wealth and life opportunities remain, though to what degree and extent is controversial. These facts raise questions about how much good capitalism has done, how capitalism might be altered or constrained to do more good, and whether another economic system would lead to better outcomes.

One trenchant critique of capitalism since Marx has been that the hypercompetition of capitalism requires business to exploit vulnerable workers, resulting in massive child labor and human

trafficking. Likewise, businesses are motivated to exploit the natural environment when it is to their benefit, at the social cost of environmental degradation. Marx also criticized capitalism for its tendency to bring about cyclical market failures, a prediction that has been borne out by devastating worldwide depressions and recessions. Although economic science has increased understanding of how government intervention can dampen these cycles, recent experience suggests that they cannot be eliminated.

Another criticism of capitalism is that it destroys traditional culture by encouraging and enabling social mobility, motivating rural to urban migration, and creating a homogenizing global culture. While capitalism clearly affects cultural evolution, whether that is for good or ill remains highly controversial. For communitarians, the destruction of traditional value is an evil in itself, while for liberals, cultures are the instruments of the individuals who comprise them, and the important question is whether capitalism coerces individuals to act in ways they would not choose for themselves.

Ann E. Cudd

See also Cultural Evolution; Heterodox Economics; Libertarianism, Political; Markets and Economic Theory; Marxist Economics; Welfare Economics; World-Systems Analysis

Further Readings

Cudd, A. E., & Holmstrom, N. (2011). *Capitalism, for and against: A feminist debate*. New York, NY: Cambridge University Press.

Held, D., & Kaya, A. (Eds.). (2007). *Global inequality: Patterns and explanations*. Cambridge, England: Polity Press.

Marx, K. (1887). *Capital* (Vol. 1). Moscow, Russia: Progress.

Sen, A. (1999). *Development as freedom*. New York, NY: Random House.

Causal Explanation, in Philosophy of Science

The truism that scientific inquiry aims not merely to *describe* phenomena but to *explain* them gives rise to one of the central questions of philosophy of science: What makes some claims about some phenomenon *count* as a scientific explanation of that phenomenon? Since at least the late 1970s, the dominant approach has endorsed this answer: Explanatory claims must provide information about the *causes* of the given phenomenon. Here, we'll look at some motivations, challenges, and open questions for this approach.

Causal accounts grew up in response to the dominant approach that preceded them: The deductive-nomological (DN) account was championed by logical empiricists such as Carl Hempel. On this latter view, we explain some phenomenon by producing a valid derivation of it (more exactly, of a sentence describing it) from true statements, at least one of which states a law of nature; the DN account adds that each such law-premise must be essential to the derivation, in that its removal renders the derivation invalid. Example: A certain pendulum has a certain period. Why? What explains this fact? Answer: The pendulum has a certain length, and there is a law that gives the period of any pendulum as a function of its length; from these two facts (but not from the first fact alone), the period of the given pendulum can be validly derived.

The DN account faced devastating counterexamples—what's more, counterexamples that seemed to uniformly point in the direction of a *causal* account as the best alternative. For example, the law that gives the period of any pendulum as a function of its length likewise gives the length as a function of the period; but for all that, we cannot explain why some pendulum has the length it does by citing its period. Evidently, what goes wrong in this example is that an *effect* is being used (together with a relevant law) to "explain" its *cause*. Again, from an effect A of some cause C, one may be able to deduce (via suitable laws) a *distinct* effect B; one does not thereby *explain* B by reference to A. (The barometer reading does not explain the subsequent storm.) Or some event C may be poised to bring about effect E if E's *actual* causes fail to do so; a mere backup such as this does not explain E, even if it can, with suitable laws, be used to derive the occurrence of E. (The victim received a fatal dose of poison but in fact died from gunshot wounds; so the poisoning does not explain his death.)

Hence the alternative, which avoids these counterexamples in the most straightforward manner: Leave laws and derivation out of the picture and insist, instead, that an explanation of some phenomenon consists in some amount of information about its causes. Still, a variety of challenges remain for any such causal account. Some can be met by

developing the account with sufficient care and flexibility. For example, scientists often seek explanations for *regularities*—which, unlike *events,* do not even seem to be the sorts of things that *have* causes. A reasonable response: We give a causal explanation of a regularity by articulating not its *causes* but rather common patterns in the causes of each of its *instances*.

Other challenges are more stubborn, and of these, the most significant is probably also the simplest: Just what *is* causation, anyway? It's no good pretending that our ordinary notion of causation is so clear and sharp that the question can be safely ignored. Consider—to cite just one of many examples testifying to its lack of clarity and precision—the controversy surrounding the question of whether "omissions" can be causes. Billy promises to water Suzy's plants and fails to do so; the plants die. Did his failure to water them *cause* their death? We might ordinarily say so, but it is far from clear, even so, that we mean the same sort of thing that we would report by saying that his trampling the plants caused their death. It thus seems an urgent question to augment a causal account of explanation with an account of causation itself.

The best contemporary approach to attempts to connect causation to *counterfactual dependence* is roughly as follows: A cause of some event E is some earlier event C where *had C not occurred, E would not have occurred*. Even this simple account fits a remarkably wide range of cases remarkably well; still, it won't do, and two reasons are especially noteworthy. First, there are cases where an event would have happened *anyway,* even had one of its causes failed to materialize; our earlier example of the poisoned assassination victim was just such a case. Second, it turns out—frustratingly if predictably—that the crucial conditional connective "had X been the case, then Y would have been the case" hides complications that themselves need to be addressed.

Very often, these conditionals submit to a fairly simple recipe for evaluation. Example: At noon, Suzy throws a rock at a window, which breaks a few moments later. What would have happened had she not thrown the rock? We answer this question by constructing a "possible world" in which, at noon, every material condition is exactly as it is in the actual world, with only the exception that Suzy is not throwing her rock (but is instead, say, standing idle); we then consider how the underlying laws of nature would have yielded an alternative future from this alternative starting point. (Presumably, in this alternative future the window does not break; hence, we may assert that if Suzy had not thrown her rock, the window would not have broken.)

A good and well-motivated recipe, as far as it goes. The trouble is that it does not go far enough for all of our scientific purposes. For in the social sciences in particular, we often seek causal explanations for phenomena, where these explanations reside in factors that cannot be intelligibly "counterfactually varied" while holding all *other* factors *fixed* in the way our recipe evidently requires. For given how intertwined and even inter*defined* different social, political, economic, and cultural factors can be, there may be no remotely realistic way to envisage *one* being different while all *others* remain as they actually are. For that reason (among others), it remains an important open question how best to develop the insight that explaining is a matter of detailing causes, so that this insight will prove as illuminating of practice in the social sciences as it already has of practice in the "hard" sciences.

Ned Hall

See also Causation, Philosophical Views of; Causation in the Social Sciences; Explanation, Theories of; Mechanism and Mechanismic Explanation

Further Readings

Hall, N. (2004). Two concepts of causation. In L. Paul, E. Hall, & J. Collins (Eds.), *Causation and counterfactuals* (pp. 225–276). Cambridge: MIT Press.

Lewis, D. (Ed.). (1986). Causal explanation. In *Philosophical papers* (Vol. 2, pp. 214–240). Oxford, England: Oxford University Press.

Salmon, W. (1998). *Causality and explanation.* New York, NY: Oxford University Press.

Strevens, M. (2011). *Depth: An account of scientific explanation.* Cambridge, MA: Harvard University Press.

Woodward, J. (2005). *Making things happen: A theory of causal explanation.* New York, NY: Oxford University Press.

Causation, Philosophical Views of

The social sciences seek causes, such as the cause of the obesity epidemic. Social scientists find causal explanations for such phenomena and use them

to predict their future development. They find and model causes to make social policy interventions to control such phenomena. What is it that the social sciences are trying to find, model, and use?

The philosophical literature on causation offers a variety of views. These have grown in sophistication since David Hume analyzed causality by appealing to effects regularly following their causes. We know now that such regularity does not always imply causation: Night follows day, but day does not cause night! And causes do not always produce their effects: Not everyone who lives in an obesogenic environment becomes obese.

This entry offers a guide to elements of the philosophical literature of interest to the social sciences. First, an important division in approaches to causality is examined: dependence or difference-making views versus production views. Then, three difference-making views are explained—probabilistic causality, the counterfactual theory, and interventionism—followed by two production views—processes and mechanisms. Finally, the relevance of these distinctions to social science is discussed.

Difference Making and Production

Ned Hall points out that many recent philosophical views fall into one of two categories:

1. *Difference-making* causes make a difference to their effects (if Jack *hadn't* grown up in an obesogenic environment, Jack *wouldn't* have been obese; or obesity rates *vary* with obesogenic factors in the environment).
2. *Production* causes are connected to effects (obesogenic factors interact with human psychological and physiological mechanisms to produce obesity).

Many take these approaches to be incompatible, but each is thought to capture something vital about causation.

Difference Making: Probabilistic Theory

Hans Reichenbach, I. J. Good, and Patrick Suppes were among the first to champion probabilistic views of causation. The simple idea is that causes raise the probability of their effects. Probability raising is a particular kind of difference making. It can be expressed formally using probability theory, as can more complex, related notions. One advantage is that this notion of causation is closely tied to science, which uses probabilistic relations extensively. Also, it can cover cases where the effects do not always follow their causes: Living in an obesogenic environment raises the probability that Jack will become obese but does not make it certain.

Unfortunately, not all causes raise the probability of their effects. Suppose Jack has a gene that strongly predisposes him to obesity, so that the probability of Jack's being obese is close to 1. Then, abundant high-calorie foods might causally contribute to Jack's obesity, without increasing the already very high probability of his obesity.

Difference Making: Counterfactual Theory

Reasoning counterfactually is common. In social science, we might claim, "If Jack had not been raised in an obesogenic environment, then he wouldn't have become obese." This is a *counterfactual conditional* because it is a contrary-to-fact conditional. A conditional is an if-then claim, and this one is counterfactual because, as a matter of fact, Jack *was* raised in an obesogenic environment. But we suppose, contrary to fact, that Jack was not raised in such an environment, and imagine what would have happened differently.

David Lewis made this kind of reasoning precise. He suggested that to assess whether a counterfactual conditional is true, we compare the actual situation with other similar "possible worlds." In similar possible worlds—such as ones where Jack's counterpart has a similar physiology and psychology but where he is not raised in an obesogenic environment—he does not become obese. Lewis then gives an account of causation in terms of counterfactuals. Effects counterfactually depend on their causes: If the cause hadn't been, then the effect wouldn't have occurred either. The account can be extended to probabilistic cases. Counterfactual dependence is the second kind of difference making.

This is a nice reasoning strategy, but Lewis claimed that possible worlds are real. There really is a world where Jack's counterpart grows up in a nonobesogenic environment. It is just not the actual world. Many philosophers reject this metaphysical claim and treat possible worlds as *fictions* we use to structure our reasoning.

Furthermore, there seem to be cases of causation without counterfactual dependence. Suppose Jack's mother is anxious and inclined to overfeed him, even at the expense of the rest of the family. Even if the family lived outside the obesogenic richer nations, Jack might still become obese. So the counterfactual "If Jack had grown up in a nonobesogenic environment, Jack would not have become obese" is not true. Even so, when Jack does grow up in the rich obesogenic world, the abundant availability of high-calorie foods is a contributing cause of Jack's obesity.

Difference Making: Intervention

James Woodward gave an account of causation that tied it closely to the kinds of experimental manipulations we use to find causes. The intuitive idea is that if we wiggle the cause, we affect the effect. If we increase the availability of high-calorie foods, we increase obesity; if we reduce the availability of high-calorie foods, we reduce obesity.

More formally, for Woodward, obesogenic factors cause obesity if the relationship between obesity and obesogenic factors is invariant under intervention on these factors, given a specific context, such as a particular age-group in a particular country. The relationship between, say, abundant high-calorie foods and obesity is invariant when it is not disrupted when we intervene to alter the system, by changing the availability of foods or slightly altering the context, such as the age-groups we assess. To "intervene" properly, we have to be able to affect the abundance of high-calorie foods without disturbing other causes of obesity, such as parental behavior. Invariance under intervention is the third kind of difference making.

When such relationships can be found, they are useful, particularly for effective policy. However, satisfying the requirements is difficult in the social sciences. Causal relationships are often sensitive to slight changes in context, so that what works for 10-year-olds might fail for 15-year-olds. And policy interventions often alter the delicate causal structure we are examining. Regulations reducing the availability of high-calorie foods to children are likely to alter parental behavior as well. For example, reassured parents may reduce their supervision of their children's diet—which may make the policy ineffective.

Production: Process

Production accounts focus on a link between cause and effect, rather than on causes making some kind of difference to their effects. For a difference-making relation to hold, it doesn't matter *how* it holds. The link between cause and effect can be left unexplained so long as a probabilistic, counterfactual, or invariant relationship between cause and effect can be detected. Production accounts focus on the *how.*

Wesley Salmon and Phil Dowe hold the view that *processes* are continuous world-lines in space-time. Causal processes are those that, when they intersect, exchange conserved quantities. This is what happens when billiard balls collide. Airplanes' shadows crossing on the ground do not exchange anything and remain unaltered after the interaction. They are therefore pseudo-processes, not causal processes.

This gives a nice account of a causal link and is the first of the two production accounts that we consider. However, it is far from clear how to apply it to anything except certain areas of physics—and certainly not to social science. If processes are important to social science, a different account is needed.

Production: Mechanism

When social scientists find a correlation (a probabilistic relationship) that they don't understand, like that between a mother's education and infant mortality, they naturally wonder whether there is a plausible *mechanism* by which a mother's education can affect infant mortality. This is another way of looking for a connection, and the second production account.

Major philosophers working on mechanisms include Peter Machamer, Lindley Darden, and Carl Craver (MDC), and Stuart Glennan. They broadly agree that mechanisms are found in three steps. For example, suppose the mechanism for a mother's education improving infant mortality is sought:

> *Phenomenon:* We want the best description of the phenomenon we can get. When does it happen? In what subpopulations? For example, just mothers' education or fathers' too? How young do the children have to be to be affected? Do you have to improve education a lot or just a little?
>
> *Parts and what they do:* Who takes children to the doctor and follows the doctor's instructions? Does

education help them understand when to consult the doctor and how to follow orders? What are the other influences on infant mortality, such as nutrition? How does maternal education affect those? Are the educated just more likely to have access to doctors and be able to afford medicine and good food anyway?

Organization: Links between access to doctors, the ability to understand instructions, and other factors like nutrition

When the whole picture is built up, and the sequence of interactions between parts is experimentally established, the mechanism has been discovered.

MDC and Glennan agree that when there is a mechanism, there is at least one causal relation. But only Glennan gives an analysis of causation in terms of mechanism: Causes are connected by mechanisms (except perhaps the causes in fundamental physics).

This view is more widely applicable than that of Salmon-Dowe and connects nicely with real scientific practice. But there is no guarantee that when a mechanism is found, an overall causal relation also exists. Mechanisms may cancel out each other's effects. When regulations limiting children's access to high-calorie foods make parents reduce their supervision, mechanisms of parental control cancel out the effect of mechanisms of regulatory control. Finally, it is often difficult to find such mechanisms, and there is no guarantee that they will always be available in social science.

Conclusion

Difference-making views and production views seem incompatible. Whatever difference-making view is used, it is possible to have causation in its absence: If there is a mechanistic link between high-calorie foods and obesity, then there may be a causal relation, even if it does not show up probabilistically, counterfactually, or in invariance under intervention. On the other hand, it seems that there can be causation without production. If nonobesity is probabilistically related to, and/or counterfactually dependent on, the absence of high-calorie foods, then this absence is a candidate cause of nonobesity. But it is difficult to see what mechanism could connect an *absence* with nonobesity.

This leaves us with a conundrum: How exactly is causation related to difference making and production? It is currently very much up in the air as to the correct response to this question.

However, even with such problems, these distinctions are useful to clarify thinking about *evidence* for causal claims. Suppose we have evidence of a correlation between obesogenic factors in the environment and obesity, across a variety of age-groups and social conditions, such as city/country and so on. This is population-level probabilistic evidence. What Jack's doctor wants to know is what to recommend *to Jack*. How should Jack alter his behavior? What past actions would have altered Jack's current situation? This is a single-case counterfactual dependence. The gap between these two concerns might be unbridgeable—or it might not. Perhaps there is evidence concerning the operation of mechanisms in Jack, or very similar individuals, which might be brought to bear. The distinctions explained here help clarify such questions.

Jon Williamson and Phyllis McKay Illari

See also Analytical Sociology and Social Mechanisms; Causal Explanation, in Philosophy of Science; Causation in the Social Sciences; Mechanism and Mechanismic Explanation; Probability; Scientific Method

Further Readings

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(Suppl. 3), 342–353.

Hall, N. (2004). Two concepts of causation. In L. Paul, E. Hall, & J. Collins (Eds.), *Causation and counterfactuals* (pp. 225–276). Cambridge: MIT Press.

Lewis, D. K. (2000). Causation as influence. *Journal of Philosophy, 97,* 182–197.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67,* 1–25.

McKay Illari, P., Russo, F., & Williamson, J. (Eds.). (2011). *Causality in the sciences.* Oxford, England: Oxford University Press.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton, NJ: Princeton University Press.

Steel, D. (2008). *Across the boundaries: Extrapolation in biology and social science.* New York, NY: Oxford University Press.

Woodward, J. (2003). *Making things happen: A theory of causal explanation.* New York, NY: Oxford University Press.

Causation in the Social Sciences

Learning the cause, or causes, of something is often important for explaining it and for designing effective plans for changing it in desired ways. Not surprisingly, then, questions of causation are usually at issue when social scientists offer explanations or policy advice. For example, economists might attempt to learn about the causes of economic recessions in order to explain why they happened when they did as well as to offer helpful suggestions to politicians about how to reduce the chance that one will occur in the near future. There are, however, a number of long-standing philosophical issues related to causation.

One group of issues concerns the concept of causation itself: A variety of conflicting theories about the nature of causation exist, and a few philosophers have even argued that the concept is bankrupt and should be dispensed with altogether. A second class of issues has to do with how knowledge about cause and effect can be acquired. It is generally agreed that correctly inferring cause-and-effect relationships in social science is very difficult, and philosophers and social scientists have offered a number of, sometimes competing, proposals about how, or whether, those difficulties may be overcome.

This entry treats these two issues in turn, discussing some interconnections between them and some implications for social science methodology.

Skeptical Challenges to Causation

One of the most fundamental questions about causation is "What is it?" We normally think of causation as something "out there" in the world. It seems an obvious and objective fact of nature that some things make other things happen. But several illustrious figures in the history of Western philosophy have argued that this commonsense idea of the objective reality of causation is an illusion. In particular, both David Hume and Bertrand Russell argued that causation, as it is normally understood, does not exist in the physical world but is instead largely a projection of the human mind onto experience.

Hume asks us to consider examples of cause and effect, for instance, one billiard ball colliding with another and sending it into the corner pocket. When we examine such examples as carefully as we can, Hume says, we only observe three basic features: (1) temporal priority (the cause happens before the effect), (2) contiguity (the cause is adjacent to the effect in space and time), and (3) constant conjunction (events like the cause are always followed by events like the effect). According to Hume, this is all causation really is in the world. This is Hume's first or "philosophical" definition of causation. But Hume points out that our ordinary concept of cause and effect involves more than just this. It includes the idea of a necessary connection between cause and effect: We think that given the cause, the effect *had to happen*. But Hume claims that, try as we might, we cannot observe any such connection between objects in the world. According to Hume, the source of our idea of necessary connection is internal rather than external. After having repeatedly observed events of Type A being followed by events of Type B, we *feel,* upon seeing a new instance of A, a strong and vivid expectation of B. Our mind then projects that expectation onto external objects. This leads to Hume's second or "natural" definition of causation, namely, that *c* is a cause of *e* if *c* is prior and contiguous to *e* and the observation of *c* "determines the mind to form" a vivid idea of *e*. Thus, Hume thinks that causation as normally understood is, to a significant extent, an illusion: We think it is "out there" in the external objects, but the most important element of it—necessary connection—is something that our mind projects on what we observe.

Bertrand Russell agreed with Hume's arguments about necessary connections not really being "out there" in the world. But Russell took skepticism about causation one step further. He argued that Hume's first, or "philosophical," definition of causation does not correspond to anything physically real either. More specifically, Russell argued that Hume's first definition of causation cannot correspond to anything in the world if modern physics is true. Russell claimed that laws of physics (here he was thinking primarily of Newtonian mechanics) do not include any reference to time; they do not say things such as "When *A* happens, *B* follows." Rather, they state relations among features of the world at a given instant. For example, Newton's law of gravitation says that the gravitational force between two objects is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.

Moreover, any empirical generalization of the form "When *A* happens, *B* follows" that could be stated by humans would have exceptions. The reason is that some event might intervene after the occurrence of *A* to prevent *B*.

Thus, Russell reasoned, if we accept Hume's first definition of causation, it follows that trying to discover cause and effect (and telling scientists that this is what they should be doing) is a waste of time, or worse. Our best option, Russell thought, is to jettison the notion of cause altogether. Instead of causation, Russell thought that it would be best to talk about the sorts of functional dependencies represented in physical laws, such as Newton's law of gravitation. Another variant of this type of view, which also appeared around the turn of the 20th century, was advocated by the statistician Karl Pearson, who argued that the notion of causation should be replaced by the (then) newly invented statistical concept of correlation.

The Causal Renaissance

It is hard to overstate the impact of Hume and Russell on subsequent discussions of causation in philosophy and statistics, and subsequently on social science. One effect was to cast a shroud of metaphysical suspicion over causation, suggesting that it is something best avoided. However, causation began to make a comeback as a respectable philosophical concept in the second half of the 20th century. In what might be regarded as the first wave of this causal renaissance, a number of philosophers proposed updated versions of Hume's first definition of causation. One example of this type of proposal asserts that *A* is a cause of *B* if *A* raises the probability of *B* even when a set of other background conditions *C* is controlled for. This concept of causation is relevant to social science, because it suggests that it would be possible to discover causes of social phenomena (e.g., recessions, political insurrections, social cohesion, etc.) through a careful examination of statistical data. However, philosophers quickly uncovered a number of difficulties in attempts to define causation in terms of probabilities. For example, suppose there are two causal pathways linking *A* to *B*: one pathway that increases the chance of *B* and the other that lowers it. Then it is possible that *A* is a cause of *B,* yet this does not raise (or lower) its probability overall. To cite one commonly given example, taking birth control pills increases the chance of thrombosis (the formation of blood clots that block blood vessels), but birth control pills also decrease the chance of pregnancy, which is a risk factor for thrombosis. Thus, despite the causal relationship between the pills and thrombosis, it is possible that taking the pills would make no difference to the probability of thrombosis. As a result of this and other objections, the idea that causation could be defined by probabilities was mostly abandoned.

The gradual demise of probabilistic definitions of causation led to a variety of alternative proposals. One line of thought focused on the mechanisms, or pathways, through which causal influence is transmitted. One of the best-known theories of this type, developed by Wesley Salmon, is founded on the concepts of *causal process* and *causal interaction*: Causal processes are objects capable of transmitting "marks," while causal interactions occur when two causal processes intersect and both are modified as a result. For example, a cue ball is a causal process that one could "mark" with momentum by hitting it with a cue stick. The cue ball could transmit this momentum to a stationary 8-ball, imparting momentum to the 8-ball and, thereby, causing it to roll into the corner pocket. One difficulty facing this theory is that it is unclear how its two key concepts can be defined without already presupposing the notion of causation. A second difficulty is that the mark transmission theory provides little help for distinguishing those chains of causal interactions that are relevant for explaining an outcome from those that are not. For instance, striking the cue ball with the stick also made a sound, which consists of vibrations in the air that also intersect with the 8-ball.

A third approach to causation emphasizes the link between causation and intervention, proposing that causal relationships differ from mere correlations in terms of indicating effective strategies for altering an outcome. Theories of this type often explicitly abandon the philosophical project of defining causation in strictly noncausal language and, instead, aim to explicate the relationship between causation and a cluster of related concepts, including intervention, explanation, and probability. James Woodward developed an approach to causation along these lines, according to which the distinctive feature of causal generalizations is that they are invariant, or continue to hold true, when subjected to interventions. Woodward conceives of an intervention as

an ideal experimental manipulation that targets a single variable without directly affecting anything else. One objection to Woodward's theory is that it is unclear how it would be applied in cases in which interventions are not possible to perform.

Causation and Social Science Methods

Each of the theories about causation described in the previous section can be helpfully thought of in relation to corresponding methods for discovering causes, methods that can be employed in social science. Probabilistic definitions of causality suggest that causes of social phenomena can be learned by carefully sifting through social statistics in just the right way. Mark transmission theories suggest that causation can be established by tracing a chain of interactions from cause to effect. In social science, this might translate into ethnographic-style analyses of social interactions, for instance, of processes through which teenagers decide to finsh high school or drop out. Finally, theories of causation that focus on intervention can be thought of in connection with experiments, either deliberately designed or naturally occurring. Moreover, methodological disputes about the necessity or relative importance of these methods in social science are not uncommon.

For example, a number of philosophers and social scientists have argued that establishing a causal relationship in social science is only possible when a mechanism has been traced from cause to effect. According to this view, for example, no amount of statistical data showing that people who regularly attend religious services tend to be happier could demonstrate a causal link between the two unless a mechanism linking them has been provided. A plausible mechanism in this case would be that regular attendance at religious services makes a person more likely to be integrated into a social support network. Without carefully documenting such a mechanism, the argument goes, we can never be sure that the correlation is not due to some unmeasured third factor that is a cause of both attending religious services and happiness. Critics of this argument respond that tracing a mechanism is not the only possible way to rule out alternative explanations.

It is possible, however, to combine at least some key ideas from separate theories of causation in ways that can illuminate connections between distinct methodological approaches. For instance, a quickly expanding interdisciplinary body of work on causation in philosophy, statistics, and computer science utilizes graphical models known as Bayesian networks to represent causation. This approach can formulate assumptions implicit in probabilistic theories of causation and can explore the extent to which those assumptions would enable causal conclusions to be inferred from statistical data. But the approach can also be used to represent interventions and hence can examine what can be learned about causation in (possibly imperfect) experiments, as well as what can be predicted about the result of an intervention from a presumed causal hypothesis.

Daniel Steel

See also Bayesianism, Recent Uses of; Causal Explanation, in Philosophy of Science; Causation, Philosophical Views of; Explanation, Theories of; Mechanism and Mechanismic Explanation; Microfoundationalism; Models in Social Science; Scientific Method

Further Readings

Hume, D. (1978). *A treatise of human nature*. Oxford, England: Oxford University Press.

Neopolitan, R. (2004). *Learning Bayesian networks*. Upper Saddle River, NJ: Prentice Hall.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). New York, NY: Cambridge University Press.

Russell, B. (1918). On the notion of cause. In *Mysticism and logic and other essays* (2nd ed., pp. 180–208). London, England: Longmans, Green.

Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.

Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam, Netherlands: North-Holland.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York, NY: Oxford University Press.

Causes Versus Reasons in Action Explanation

The debate whether human action is the result of, and can be explained by appeal to, reasons as opposed to causes has been of central importance both to the

philosophy of action and to the philosophy of the social sciences. The debate was particularly significant to the latter, especially because it was thought that if reasons for actions are not causes, then the social sciences cannot be assimilated to natural sciences and understanding intentional behavior is distinct from explanation of physical phenomena.

Causation

Philosophers theorizing about causation have by and large followed David Hume in construing it as a relation between events in succession such that, given the types of events they are, once the first occurs the second is bound to follow. Most today reject the so-called Humean regularity thesis of causal relations and Hume's claim that this causal necessity is a "determination of the mind," a human "habit," and not found "in the objects themselves." The dominant picture is that general laws governing events of certain types "subsume" causal relations between particular instances of those types. It is also part of this view that causal relations and the laws that govern them are found in nature, independently of our particular ways of identifying them, and that we discover these relations through experience and experiment.

There are good reasons for rejecting at least some of the orthodoxy just outlined about the "nature" of causation. It seems to depend on the idea that the nouns *cause, causation,* or *causality* name something general and abstract, and it is the metaphysician's business to enquire into its nature. A more satisfying but underappreciated view is that "causation" names nothing: It is an umbrella or "concept" term that "collects" a variety of uses of expressions that resemble one another in disparate ways. Rather than speculating about the nature of the object or the property allegedly named by the abstract noun, it is the philosopher's job to trace the contours and details of these uses and compare and contrast them with neighboring concepts, with which they abut, crisscross, and overlap. This kind of investigation reveals that, like all expressions that are collected by general concepts that have their home in our everyday practices, causal terms change their inflections from one circumstance to another. Therefore, we can and should talk about "central" cases: those that are at some remove from the center and those located toward the periphery. These can be compared and contrasted with the multifarious uses of expressions we collect under the umbrella term *reasons for action.*

In its central uses, the concept of causation is inextricably bound with explanation and prediction. Construing causation as an "external" relation between events—a relation "in the world" that occurs independently of our ways of conceptualizing or describing it—has the effect of severing causation from causal explanation, and the metaphysical analysis is thus allowed to float free from the role the concept plays in our lives, with most unsatisfactory results. Concentrating on the genuine role the concept plays will reveal that our acquaintance is not with a general relation, namely, that of "causation," *in itself,* that holds between particular events and even less with the natural laws that subsume them.

Although necessity enters into our understanding of causation in central cases—unnecessitated causes are the exception and not the norm—it is not as a mysterious metaphysical glue holding together the events themselves or as an entailing power in the universal laws that subsume them.

That is, *for us and in the particular circumstances we are considering,* and not in other "conceivable" ones, there is no more scope for avoidance in, say, getting wet when swimming than for failure in constructing a heptagon with a compass and straightedge. Both the natural and the mathematical impossibilities are equipollent for us.

Primary cases of causation, then, seem to involve a circumstance-dependent, nonidealized notion of necessity and the notion of one object or event bringing about or having an impact on another. In the central cases, we learn of these interactions through experience and experiment.

Explaining

Does this short description of causation tell us what we need to know about explanation, generally construed, as some philosophers seem to think? In order to see why the answer is negative, consider the following questions and answers.

1. Why do we need 120 wine glasses? Because we have 30 guests and 4 tastings.
2. Why did he become so angry when we offered to help him with the bill? Because he is proud.

3. What makes you think Christmas falls on a Sunday? Because New Year's Day is on a Sunday and they are a week apart.
4. What makes him a bachelor? Because he is unmarried.
5. Why is this animal a mammal? Because it is a whale.

In contrast, consider the following:

6. Why was he right to concede the match? Because there was no way he could win.
7. Why add the olive oil so slowly to the beaten egg? Because otherwise it will curdle.

We can imagine situations in which the reasons given in each of these answers will satisfy us by making the phenomenon to be explained unmysterious. The explanations given are backed by or grounded in (either explicitly or implicitly) diverse considerations: the first, on a mathematical calculation; the second, on a trait of character; the third, on the conventions set out in a calendar; the fourth, on the meaning of a word; the fifth, on a biological or zoological taxonomy; the sixth, on the rules of a game; and the seventh, on past experience or to an established regularity.

But notice how few of these "becausal" explanations we would categorize—without rewriting them—as *causal* explanations. It would sound odd to say, for example, that the multiplication of 30 guests times 4 tastings caused the requirement for 120 glasses or that the man's pride caused his becoming upset. Instead, the multiplication provides justification (or proof) for the suggested number of glasses; getting upset when others offer to help is one of the manifestations, or is indicative, of a man's pride.

Nor would we be tempted to claim that New Year's falling on Sunday causes Christmas to fall on the same day or that being a whale causes it to be a mammal. Rather, the calendar determines the day on which each holiday falls, and biological or zoological taxonomy grounds the claim that whales are mammals. An (unidealized) element of necessity is present in some but not all of these cases, but in none do we have the idea of one object or event causing another.

The first six cases are different in kind from the explanation for why the oil is added slowly to the egg mixture. The reason given in this case—"because otherwise it will curdle"—does suggest a causal claim that grounds the explanation, namely, that adding oil too quickly to whipped egg yolks causes the mixture to curdle. The grounds for this claim involve past experience, or trial and error, with two substances that interact in different ways depending on how they are combined.

This shows that not all explanations are causal and only some have causal backings.

Reasons Versus Causes in Action Explanation

What about explanations of action? Example 6 is about an agent's action: that of conceding the match. Example 7 is about a rule (in this case, a recipe for mayonnaise) that governs an action: Oil should be added only very slowly to the egg yolks.

The other examples can be set out explicitly as explanations of action. Why did you put 120 glasses on the table? Why did he slam his hand on the table when we offered to pay the bill? Why did she tell them that Christmas is on a Sunday, that he is a bachelor, or that the whale is a mammal?

Does rewriting them thus require a different answer from the ones given? It depends on what we are looking for, and this, of course, depends on what is puzzling us. The very same answers will satisfy us about a person's *reasons* for calling a whale a mammal, for serving 120 glasses of wine, for pouring the oil very slowly into the egg mixture, and so on, if we are concerned about what justifies the claim or action: why someone in the same situation *should* do the same.

Ludwig Wittgenstein, John Wisdom, Gilbert Ryle, Elizabeth Anscombe, and others were keen to point out the circumstance dependence of explanation: what counts as a successful one depends on whom it is for and how this understanding informs one's abilities and practices. Explanations come to an end somewhere, Wittgenstein reminds us. Knowing a law of nature or an inference rule, Ryle suggests, depends on knowing how to use this law in making explanations or predictions or in implementing the rule in deducing a conclusion from a set of premises. Explanations, as it were, do not sit in some metaphysical heaven to be intuited by the mind: They are woven into the fabric of our daily practices. Even so, these philosophers distinguished between different kinds of explanation. Ryle accepts

a certain picture suggested by the use of "cause," in which causal laws are established by experiment and the relata that enter into causal relations are independently observable and describable. This he contrasts with a phenomenon that is partly constitutive of another, as the headside of a penny is part of what makes it the penny it is, but is not a separately existing agency that causes it to be so. For Wittgenstein, a crucial contrast is between what is discovered on the basis of experiment versus the rules we lay down. Anscombe famously traced the logical or grammatical ties between the concepts of intentions, motive, reason, and voluntariness, contrasting the kind of explanation facilitated with that given by citing "mental causes."

Today, however, philosophers tend to see causes and even causal *explanation* as quite independent of human contingencies and practices. Objectivity and discovery are the bywords for scientific investigation, which, many hold, seek out natural laws that allow us to explain events in terms of their causes by construing these relations as instances of more general patterns in (mind-independent) nature. Reasons, by contrast, provide explanations that allow us to see, in the paradigm case, human action as making sense based on how the agent conceives it, as well as (how she conceives) her goals, desires, beliefs, and values. (Teleological explanation, in which natural events are seen as arising from the purposes of God, Mother Nature, or evolution, sits in between this too stark division.)

If causation falls within the province of the natural sciences and if the "special sciences," such as psychology, sociology, or anthropology, are to aspire to the same measurable, repeatable, and in this sense objective standards, then what seems to be a matter of subjectivity and interpretation needs to be reconciled with the opposite aims of the natural sciences. Donald Davidson, though inveighing against the idea that there are natural laws of psychology, was nonetheless responsible for giving hope to many that there could be a science of the mind. He famously challenged the positions of Wittgenstein, Anscombe, A. I. Melden, and others by arguing that for a true explanation of an action, we need to know what events *caused* it as well as the agent's conception, values, goals, and so on, that make the action *reasonable,* at least from her point of view.

Anscombe's insightful contribution to action theory was to point out, among other things, that a human action is *intentional* or that someone acts with an *intention to act* when the question "Why did she do it?" can be meaningfully answered only by citing her *reasons* for action, and these, for Anscombe, are not causes. Furthermore, she pioneered the important idea that an action is intentional *under a description* that cites the actor's reasons or intention, while this is perfectly compatible with the same action being seen—because, for example, it is constituted by bodily movements, such as an arm rising—as a physical event *under a different description.* But the latter will not amount to an explanation of an intentional action. In addition, others, like Melden, for example, argued that intentions or reasons to act are not Humean causes, for the latter involve, as we have seen at the start, a model of causation as an external relation holding between two distinct events, whereas, by contrast, intentions are not *logically* distinct from the resulting actions—they (intending and acting) are not two logically distinct events. Citing an intention requires expressing it as "the intention to do x," where "x" describes the action—that is, once we describe an intention, we already refer to the action—thus violating the standard Humean model of causation requiring separateness of event-descriptions.

To those working in the philosophy of the social sciences, this account of action whereby reasons to act were not Humean causes provided an additional argument in favor of the nonassimilation of the study of the social world to the study of the physical world as carried out by the natural sciences. Action as meaningful behavior requires understanding the reasons behind it, not a causal explanation. Parallel to this, Peter Winch's influential thesis about rule following reinforced this anticausal stance.

Davidson, however, challenged the thesis that reasons are not causes and argued that a reconstructed rationalization of the kind he himself favored (in which reasons are "reduced" to beliefs, desires, and weighted value judgments) is still not enough to explain the agent's action, for it fails to give us *the* reason that explains the action: the reason, that is, *for which* the agent acts. His argument is that more than one rational reconstruction is possible, any one of which will provide an "anemic justification" for the action. We can, he continues, imagine cases in which the agent acts because of one of these reasons instead of another. Short of a better story, he suggests, we may as well construe the connective "because" in explanations citing reasons as signifying a *causal* relation.

Davidson himself construed causation in the Neo-Humean way—outlined at the beginning of this entry—namely, as a relation between "real-world" events, which occur independently of our way of describing or conceptualizing them. His argument that reasons are causes became extraordinarily influential in supporting a view of the mind that was also gaining momentum independently: the view that "mental states" (which, he argued, though nonreducible are nonetheless identical to tokens of or particular physical events) are individuated or identified as the kind of mental state they are by virtue of their causal role in mediating between stimulus or input, behavior or output, and other mental states. (This is the view in philosophy of mind called *functionalism,* married with *token physicalism.*) This view itself was given support from the same implicit assumption about language described earlier, that mental concepts name something—to wit, mental states, events, or properties. More precisely, Davidson held that mental verbs such as "believes," which have as their subject a pronoun or proper noun and are followed by an accusative noun or a "that . . ." clause—such as "John believes that it's time to decant the wine"—signify something: John's belief. (This in turn is construed as, e.g., "John's hat" might be—as naming a property of John, but since, unlike his hat, it is nonphysical under its description as a "belief," it becomes a property about the nature of which it is the philosopher's job to enquire). If reasons involve a rational reconstruction of "content-bearing" or "propositional-attitude" states, such as beliefs and desires and these states themselves are (identical to) physical states in the brain, then the philosopher becomes additionally burdened with the puzzle of how something in the brain could bear representational "content"—in other words, be "about" something.

As a result of his influential arguments that reasons are causes, today's orthodoxy is that a rational reconstruction of an agent's action can be given by adverting to "cognitive" and "conative" states: colloquially, her beliefs and her desires or other "pro-attitudes." The explanation why the person put 120 wine glasses on the table, on this construal, would provide us with information about what she wanted to achieve by her action (in this case, to provide enough glasses for the number of tastings) and what she believed would be necessary to meet that goal (that 3 tastings multiplied by 40 guests requires 120 glasses).

There are, however, a number of problems with the standard view. Note, first, how this reconstruction changes the original example, which is plausibly paraphrased as asking the right with which the woman put 120 wine glasses on the table. An answer to this question is not given by asking after the agent's beliefs, because these may be false. Indeed, it seems clear that we would, in the normal case, retreat from the "facts" to the agent's beliefs when her action cannot be given the normal justification (e.g., Why did she put out 80 glasses of wine? Because she believed, wrongly, that there would only be 2 tastings). Citing a person's beliefs and desires as her reasons for performing a particular action does not answer the question whether, in those circumstances, she was right or justified in doing so, or why someone else in her situation should do the same. And this is, as we have seen, one of the things we might be wondering when enquiring into someone's reasons for acting.

Second, if the action is based on a miscalculation, misinformation, wishful thinking, or something not rationalizable in the full-blooded sense, then the anemic sort of justification obtained by pinpointing what the agent mistakenly believed or perversely described is not given by looking at the causal mechanism of her mind/brain; it is given by a rational reconstruction that takes into account the less-than-ideal circumstances ("She thought there would only be 2 tastings" or "She forgot how to multiply" or "There were only 80 glasses available" or "She wanted to ruin her husband's party"). If these are not obvious, we may find the answer by asking her. Thus, Anscombe's proposed criterion, what the agent would say in response to the question "Why did you do this?" or "What were you thinking?" is also something we may mean when we ask about someone's reasons for acting. The reasons for which she acts become, on this construal, the ones she would give in answer to the question.

Third, construing reasons as causes in the primary sense, as one thing or event having an impact on another, puts reasons, as it were, in the wrong logical/grammatical/ontological category: Questions we ought to be able to ask about events (e.g., When did it occur? How long did it last?) make no sense in many of the central cases in which we enquire into a person's reasons for acting. Fourth, nor is there the element of necessity between reasons and actions that seems to attend causal claims: even the nonidealized kind. Fifth, nor is the relation between reason

and the action it justifies the one that we learn about from experience and experiment.

Furthermore, sixth, construing reasons as causes does not solve the "mind–body" problem: It resurrects it. It creates a mystery about the nature of mental stuff such that it can affect other stuff, whether mental or otherwise. Seventh, construing reasons as the causal antecedents of an agent's action introduces the problem of self-knowledge: How could an agent possibly know what her reasons are if they are at bottom physical (neurophysiological) processes?

Of course, agents cause things: The captain caused the shipwreck by showing off, and in showing off, he navigated too close to the shore. One might say that his bravado caused the shipwreck, or perhaps, to be more candid we might say that the shipwreck occurred because of it. In saying the former, we would be speaking relatively loosely and would not commit ourselves to the idea that after a thorough investigation of things in the world having an impact on one another—navigation instruments, ship mechanisms, heavy metal, and rock outcrops—there will be at least one more thing that had an impact, the captain's bravado, about the nature of which it is the philosopher's job to enquire. But unless the shipwreck was intentional on his part, he did not have a reason for causing it. Even if he did—he wanted to destroy the cruise liner in order to wreak havoc on the company and did not care if lives were lost in the process—reasons are not "things" or "processes" either; they are not to be found among the things, like the ship and the rock that played a causal role in the shipwreck.

Julia Tanney

See also Action, Philosophical Theory of; Agency; Causation, Philosophical Views of; Causation in the Social Sciences; Explanation Versus Understanding; Laws of Nature; Laws Versus Teleology; Mind–Body Relation; *Naturwissenschaften* Versus *Geisteswissenschaften*; Rule Following

Further Readings

Anscombe, G. E. M. (1957). *Intention.* Oxford, England: Blackwell.

Anscombe, G. E. M. (1979). Under a description. *Noûs, 13,* 219–233.

Anscombe, G. E. M. (2005). The causation of action. In L. Gormally & M. Geach (Eds.), *Human life, action, and ethics: Essays by G. E. M. Anscombe* (pp. 89–108). Exeter, England: Imprint Academic.

Davidson, D. (1980). *Essays on actions and events.* Oxford, England: Clarendon Press.

D'Oro, G. (Ed.). (in press). *Reasons and causes: Causalism and non-causalism in the philosophy of action* (History of Analytic Philosophy Series). London, England: Palgrave Macmillan.

Holland, R. F. (1980). The link between cause and effect. In *Against empiricism: On education, epistemology, and value* (pp. 210–226). New York, NY: Barnes & Noble.

Hume, D. (1896). *A treatise of human nature* (with an analytical index by L. A. Selby-Bigge, Ed.). Oxford, England: Clarendon Press. (Original work published 1740)

Hume, D. (1910). *An enquiry concerning human understanding* (Harvard Classics Vol. 37). New York, NY: P. F. Collier. (Original work published 1748)

Melden, A. I. (1961). *Free action.* London, England: Routledge & Kegan Paul.

O'Connor, T., & Sandis, C. (Eds.). (2009). *A companion to the philosophy of action.* Oxford, England: Blackwell.

Ryle, G. (2009). *Collected papers* (Vol. 2). London, England: Routledge.

Tanney, J. (2013). *Rules, reason, and self-knowledge.* Cambridge, MA: Harvard University Press.

Wisdom, J. (1991). *Proof and explanation* (The Virginia Lectures; S. F. Barker, Ed.). Lanham, MD: University Press of America. (Original work published 1957)

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Oxford, England: Blackwell.

Wittgenstein, L. (1975). *Wittgenstein's lectures on the foundations of mathematics: Cambridge 1939* (C. Diamond, Ed.). Chicago, IL: University of Chicago Press.

Chicago School (Economics)

The Chicago school of economics was a post–World War II group of economists, often associated with the University of Chicago, who shared a commitment to using the analytical tools of price theory to test economic policies against their empirical outcomes. Chicago economists, however, did not develop increasingly complex mathematical models or models with more realistic assumptions to predict policy outcomes. Instead, they used analytically simple models based on the basic principles of the market performance of rational individuals to identify how various government interventions in free markets created inefficiencies in the allocation of resources across the economy.

Arguing that the basic principles of economic rationality reliably predict the outcomes of individuals' actions in market settings, Chicago economists focused their attention on the structure of opportunity costs and policy-created constraints that surround economic actors, rather than on changes in tastes, values, and preferences. Often, the Chicago school's examination of a policy defied the conventional wisdom of economists, other social scientists, and legal scholars and showed the unintended (and usually negative) consequences of the policy. A classic example of the Chicago approach is the argument against minimum-wage laws. Regardless of its good intentions, legislation that creates a floor below which wages cannot fall ends up hurting the people it was created to help. Higher wage costs lead employers to reorganize their workflow and substitute capital investments for labor in order to make fewer workers more productive. The lowest productive workers would become perpetually unemployed under such legislation and might drop out of the workforce eventually.

By the 1960s, Chicago economists realized that their logic could be extended into other policy arenas. Sociological issues like discrimination, crime, urban problems, and education came under the lens of the Chicago approach, as did public and bureaucratic choices. The Chicago school spearheaded the movement of "rational choice" theory into sociology and political science. And in public sector analysis, they were closely allied with public choice scholars. Inroads into legal theory began with the transaction cost approach to economic organization and its extension into the examination of the social costs of negative externalities. Questioning the traditional argument against concentrations of market power, Chicago law and economics researchers, as well as its industrial organization theorists, argued that market dominance was fleeting unless supported by government action and regulation. Courts and legislatures should be less concerned about the market actions of corporations than previous industrial organization theorists had suggested. In most cases, activities previously considered anticompetitive were actually responses to consumer demand and transaction costs. In response to the Chicago approach, mergers, acquisitions, long-term contracts, resale price maintenance agreements, price discrimination, bundling of services, and other activities have come to be seen as legitimate activities of firms in open markets.

The Chicago school is best known in the media for its role in monetary and financial theory. Both monetarism and the theory of efficient financial markets were built upon the same basis of applied price theory as Chicago's other policy evaluations. Monetarists argued that in an economy featuring a large degree of economic freedom, the best long-term monetary strategy was a rule dictating a rate of money supply growth approximately equal to the long-term, real economic growth rate. Efficient market theorists argued that financial markets that were free and diversified would efficiently process all available market information without regulatory control. New firms and new financial products would enter the financial markets to provide participants with better means of capturing available returns, as they do in consumer markets. The monetary stability of the 1990s and early 2000s, as well as the concurrent expansion of the financial markets and diversification of financial derivatives, is often attributed to the theories of the Chicago school.

Three types of criticisms have been brought against the Chicago school. Each type of criticism, while focused on the school's policy evaluations, bears the marks of an underlying philosophical difference with the school's approach. The first and most common criticism is the charge that ideology trumps scientific inquiry in the school. The criticism came to the forefront of public attention when policymakers in the United Kingdom, Latin America, Israel, post-Communist Russia, and elsewhere turned to the Chicago school for insights on liberalizing their economies. Because the Chicago approach almost always returns the policy advice to lessen social control and expand market freedom, the critics have ample opportunity to wonder if the inquiry leads to the conclusion or the conclusion drives the inquiry.

The second type of criticism focuses on the school's use of simple price theory models to evaluate policy outcomes. The argument here can take several directions. Many have argued that economics advances by developing models with more nuanced assumptions, in terms of either their realism or their generalizability. Another common argument is that a social scientist cannot ignore changes in values and preferences, and in social influence on these changes.

The final type of criticism is the most fundamental philosophically: the school's dependence on the combination of an instrumental conception of rationality (homo economicus—the rational economic

person) and an individualistic conception of society. Any social scientific approach that views humans as more than rational calculators or society as more than a social contract will find the Chicago school problematic at some level.

And yet the Chicago school was remarkably successful in the latter half of the 20th century. Even when economists today disagree with some aspects of the school's findings, their underlying approach to policy evaluation is deeply influenced by the school's legacy.

Ross B. Emmett

See also Homo Economicus; Individualism, Methodological; Policy Applications of the Social Sciences; Rational Expectations; Rationality and Social Explanation

Further Readings

Becker, G. S. (1976). *The economic approach to human behavior.* Chicago, IL: University of Chicago Press.

Coase, R. H. (1960). The problem of social cost. *Journal of Law & Economics, 3,* 1–44.

Emmett, R. (Ed.). (2010). *The Elgar companion to the Chicago school of economics.* Cheltenham, England: Edward Elgar.

Friedman, M. (1953). The methodology of positive economics. In *Essays in positive economics* (pp. 3–43). Chicago, IL: University of Chicago Press.

Reder, M. W. (1982). Chicago economics: Permanence and change. *Journal of Economic Literature, 20,* 1–38.

Classical Computationalism, Connectionism, and Computational Neuroscience

Computationalism is the view that cognitive capacities have a computational explanation or, somewhat more strongly, that cognition is (a kind of) computation. For simplicity, these two formulations will be used interchangeably. Most cognitive scientists endorse some version of computationalism. Thus, when cognitive scientists propose an explanation of a cognitive capacity, the explanation typically involves computations that result in the cognitive capacity. But cognitive scientists differ on how they apply the notion of computation to the study of cognition. Their three main research traditions are classical computationalism, connectionism, and computational neuroscience.

This entry introduces the notions associated with cognition construed as computation and reviews in detail the three traditions.

Some Notions of Computation

Computationalism is usually introduced as an empirical hypothesis that can be disconfirmed. Whether computationalism has empirical bite depends on how we construe the notion of computation: The more inclusive a notion of computation, the weaker the version of computationalism formulated in its terms.

At one end of the continuum, some notions of computation are so loose that they encompass virtually everything. For instance, if computation is construed as the production of outputs from inputs and if any state of a system qualifies as an input (or output), then every process is a computation. Sometimes, computation is construed as information processing, which is somewhat more stringent, yet the resulting version of computationalism is quite weak. There is little doubt that organisms gather and process information about their environment.

The best computational theories appeal to a well-defined kind of computation. Historically, the most important version of computationalism appeals to *digital* computation—the kind of computation performed by digital computers. Other versions appeal to *analog* computation or computation *in a generic sense,* which encompasses both digital and analog computation, among others, as species.

Three Research Traditions: Classical Computationalism, Connectionism, and Computational Neuroscience

Computation: Origins

The view that thinking has something to do with computation may be found in the works of some modern materialists, such as Thomas Hobbes in the 17th century. But computationalism properly so called could not begin in earnest until a number of 20th-century logicians (most notably Alonzo Church, Kurt Gödel, Stephen Kleene, Emil Post, and especially Alan Turing) laid the foundations for the mathematical theory of computation.

Turing analyzed computation in terms of what are now called "Turing machines"—a kind of simple processor operating on an unbounded tape.

The tape is divided into squares, which the processor can read and write on. The processor moves along the tape, reading and writing on one square at a time depending on what is already on the square as well as on the rules that govern the processor's behavior. The rules state what to write on the tape and where to move next, depending on what is on the tape as well as which of the finitely many states the processor is in.

Turing argued convincingly that any function that can be computed by following an algorithm (i.e., an unambiguous list of instructions operating on discrete symbols) can be computed by a Turing machine. Church offered a similar proposal in terms of general recursive functions, and it turns out that a function is general recursive if and only if it can be computed by a Turing machine. Given this extensional equivalence between Turing machines and general recursive functions, the thesis that any algorithmically computable function is computable by some Turing machine (or, equivalently, is general recursive) is now known as the *Church-Turing thesis*.

Turing made two other relevant contributions. First, he showed how to construct *universal* Turing machines. These are Turing machines that can mimic any other Turing machine by encoding the rules that govern the other machine as instructions, storing the instructions on a portion of their tape, and then using the encoded instructions to determine their behavior on the input data. Notice that ordinary digital computers, although they have more complex components than universal Turing machines, are universal in the same sense (up to their memory limitations). That is, digital computers can compute any function computable by a Turing machine until they run out of memory.

Second, Turing showed that the vast majority of functions whose domain is denumerable (e.g., functions of strings of symbols or of natural numbers) are actually *not* computable by Turing machines. These ideas can be put together as follows: Assuming the Church-Turing thesis, a universal digital computer can compute any function computable by algorithm, although the sum total of these Turing-computable functions is a tiny subset of all the functions whose domain is denumerable.

Modern Computationalism and Neural Networks

Modern computationalism began when Warren McCulloch and Walter Pitts connected three things: (1) Turing's work on computation, (2) the explanation of cognitive capacities, and (3) the mathematical study of neural networks. Neural networks are sets of connected signal-processing elements ("neurons"). Typically, they have elements that receive inputs from the environment (input elements), elements that yield outputs to the environment (output elements), and elements that communicate only with other elements in the system (hidden elements). Each element receives input signals and delivers output signals as a function of its input and current state. As a result of their elements' activities and organization, neural networks turn the input received by their input elements into the output produced by their output elements. A neural network may be either a concrete physical system or an abstract mathematical system. An abstract neural network may be used to model another system (such as a network of actual neurons) to some degree of approximation.

The mathematical study of neural networks using biophysical techniques began around the 1930s. Before McCulloch and Pitts, though, no one had suggested that neural networks have something to do with computation. McCulloch and Pitts defined networks that operate on sequences of discrete inputs in discrete time, argued that they are a useful idealization of what is found in the nervous system, and concluded that the activity of their networks explains cognitive phenomena. McCulloch and Pitts also pointed out that their networks can perform computations like those of Turing machines. More precisely, McCulloch-Pitts networks are computationally equivalent to Turing machines without tape or finite-state automata. Modern digital computers are a kind of McCulloch-Pitts neural network. Digital computers are sets of logic gates—digital signal-processing elements equivalent to McCulloch-Pitts neurons—connected to form a specific architecture.

McCulloch and Pitts's account of cognition contains three important aspects: (1) an analogy between neural processes and digital computations, (2) the use of mathematically defined neural networks as models, and (3) an appeal to neurophysiological evidence to support their neural network models. After McCulloch and Pitts, many others linked computation and cognition, though they often abandoned one or more aspects of McCulloch and Pitts's theory. Computationalism evolved into three main traditions, each emphasizing a different aspect of McCulloch and Pitts's account.

The Three Traditions: Classical Computationalism, Connectionism, and Computational Neuroscience

One tradition, sometimes called *classical computationalism,* emphasizes the analogy between cognitive systems and digital computers while downplaying the relevance of neuroscience to the theory of cognition. When researchers in this tradition offer computational models of a cognitive capacity, the models take the form of computer programs for producing the capacity in question. One strength of the classicist tradition lies in programming computers to exhibit higher cognitive capacities such as problem solving, language processing, and language-based inference.

A second tradition, most closely associated with the term *connectionism* (although this label can be misleading; see below), downplays the analogy between cognitive systems and digital computers in favor of computational explanations of cognition that are "neurally inspired." When researchers in this tradition offer computational models of a cognitive capacity, the models take the form of neural networks for producing the capacity in question. Such models are primarily constrained by psychological data, as opposed to neurophysiological and neuroanatomical data. One strength of the connectionist tradition lies in designing artificial neural networks that exhibit cognitive capacities, such as perception, motor control, learning, and implicit memory.

A third tradition is most closely associated with the term *computational neuroscience,* which is one aspect of theoretical neuroscience. Computational neuroscience downplays the analogy between cognitive systems and digital computers even more than the connectionist tradition. Neurocomputational models aim to describe actual neural systems, such as (parts of) the hippocampus, cerebellum, or cortex, and are constrained by neurophysiological and neuroanatomical data in addition to psychological data. It turns out that McCulloch-Pitts networks and many of their "connectionist" descendants are relatively unfaithful to the details of neural activity, whereas other types of neural networks are more biologically realistic. Computational neuroscience offers models of how real neural systems may exhibit cognitive capacities, especially perception, motor control, learning, and implicit memory.

Although the three traditions just outlined are in competition with one another to some extent, there is also some fuzziness at their borders. Some cognitive scientists propose hybrid theories that combine explanatory resources drawn from both the classicist and the connectionist traditions. In addition, biological realism comes in degrees, so there is no sharp divide between connectionist and neurocomputational models.

Understanding the Three Traditions

The debate between classicists and connectionists has been somewhat confusing. Different authors employ different notions of computation, which vary in both their degree of precision and their inclusiveness. Specifically, some authors use the term *computation* only for classical computation—at a minimum, algorithmic digital computation over language-like structures—and conclude that (nonclassicist) connectionism falls outside computationalism. By contrast, other authors use a broader notion of computation, thus including connectionism within computationalism. But even after we factor out differences in notions of computation, further confusions lie in the wings.

Classical computationalism and connectionism are often described as being at odds with one another, because classical computationalism is committed to the idea that the vehicles of digital computation are language-like structures and, it is assumed, is also committed to the idea of autonomy from neuroscience—two theses flatly denied by many prominent connectionists. But many connectionists also model and explain cognition using neural networks that perform computations defined over digital structures, so perhaps they should be counted among the digital computationalists.

Furthermore, both classicists and connectionists tend to ignore computational neuroscientists, who in turn tend to ignore both classical computationalism and connectionism. Computational neuroscientists often operate with their own mathematical tools, without committing themselves to a particular notion of computation. To make matters worse, some connectionists and computational neuroscientists reject digital computationalism—they maintain that their neural networks, while explaining behavior, do not perform digital computations.

In addition, the very origin of digital computationalism calls into question the commitment to autonomy from neuroscience. McCulloch and Pitts

initially introduced digital computationalism as a theory of the brain, and some form of computationalism or other is now a working assumption of many neuroscientists.

A further wrinkle in this debate derives from the ambiguity of the term *connectionism*. In its original sense, connectionism says that behavior is explained by the changing "connections" between stimuli and responses, which are biologically mediated by the changing connections between neurons. This original connectionism is related to behaviorist associationism, according to which, behavior is explained by the association between stimuli and responses. Associationist connectionism adds a biological mechanism to explain the associations: the mechanism of changing connections between neurons.

But contemporary connectionism is a more general thesis than associationist connectionism. In its most general form, contemporary connectionism, like computational neuroscience, simply says that cognition is explained (at some level) by neural network activity. This is a truism, or at least it should be. The brain is the organ of cognition, the cells that perform cognitive functions are (mostly) neurons, and neurons perform their cognitive labor by organizing themselves in networks. Neural activity is computation at least in a generic sense. Even digital computers are just one special kind of neural network. So even classicists, whose theory is most closely inspired by digital computers, are committed to connectionism in its general sense.

The relationship between connectionist and neurocomputational approaches on the one hand and associationism on the other turns on a distinction between strong and weak associationism. Strong associationism maintains that association is the only legitimate explanatory construct in a theory of cognition. Weak associationism maintains that association is a legitimate explanatory construct along with other constructs such as the innate structure of neural systems.

To be sure, some connectionists profess strong associationism. But that is beside the point, because connectionism per se is consistent with weak associationism or even the complete rejection of associationism. Some connectionist models do not rely on association at all—a prominent example being the work of McCulloch and Pitts. Weak associationism is consistent with many theories of cognition, including classicism. A vivid illustration is Alan Turing's early proposal to train associative neural networks to acquire the architectural structure of a universal computing machine. In Turing's proposal, association may explain how a network acquires the capacity for universal computation (or an approximation thereof), while the capacity for universal computation may explain any number of other cognitive phenomena.

Although many of today's connectionists and computational neuroscientists emphasize the explanatory role of association, many of them also combine association with other explanatory constructs, as per weak associationism. What remains to be determined is which neural networks, organized in what way, actually explain cognition and which role association, and other explanatory constructs, should play in a theory of cognition.

Yet another source of confusion is that classical computationalism, connectionism, and computational neuroscience tend to offer explanations at different mechanistic levels. Specifically, classicists tend to offer explanations in terms of rules and representations, without detailing the neural mechanisms by which the representations are implemented and processed; connectionists tend to offer explanations in terms of highly abstract neural networks, which do not necessarily represent networks of actual neurons (in fact, a processing element in a connectionist network may represent an entire brain area rather than an actual neuron); finally, computational neuroscientists tend to offer explanations in terms of mathematical models that represent concrete neural networks based on neurophysiological evidence. Explanations at different mechanistic levels are not necessarily in conflict with each other, but they do need to be integrated to describe a multilevel mechanism. Integrating explanations at different levels into a unified multilevel mechanistic picture may require revisions in the original explanations themselves.

Different parties in the dispute between classical computationalism, connectionism, and computational neuroscience may offer different accounts of how the different levels relate to one another. One traditional view is that computational explanations do not describe mechanisms. Instead, computational and mechanistic explanations are independent. This suggests a division of labor: Computations are the domain of psychologists, while the implementing

neural mechanisms are the business of neuroscientists. According to this picture, the role of connectionists and computational neuroscientists is to discover how neural mechanisms implement the computations postulated by (classicist) psychologists.

This traditional view has been criticized as unfaithful to scientific practices. It's been pointed out that (a) both psychologists and neuroscientists offer computational explanations; (b) far from being independent, different levels of explanation constrain one another; and (c) both computational explanations and mechanistic explanations can be given at different levels.

One alternative to the traditional view is that connectionist or neurocomputational explanations simply replace classicist ones. Perhaps some connectionist computations approximate classical ones. In any case, some authors maintain that classicist constructs, such as program execution, play no causal role in cognition and will be eliminated from cognitive science.

A more neutral account of the relation between explanations at different levels is provided by the mechanistic account of computation. According to the mechanistic account, computational explanation is just one type of mechanistic explanation. Mechanistic explanations provide components with such properties and organization that they produce the phenomenon. Computational explanation, then, is explanation in terms of *computing* mechanisms and components—mechanisms and components that perform computations. Mechanistic explanations come with many levels of mechanisms, where each level is constituted by its components and the way they are organized. If a mechanistic level produces its behavior by the action of computing components, it counts as a computational level. Thus, a mechanism may contain zero, one, or many computational levels, depending on what components it has and what they do. Which types of computation are performed at each level is an open empirical question to be answered by studying cognition and the nervous system at all levels of organization.

Conclusion

Computationalism is here to stay. We have every reason to suppose that cognitive capacities have computational explanations, at least in a generic sense. Moreover, everyone is (or should be) a connectionist or computational neuroscientist, at least in the general sense of embracing neural computation. Nonetheless, much work remains to be done.

The computational study of cognition will require that we integrate different mechanistic levels into a unified, multilevel explanation of cognition. We also need to characterize the specific computations on which cognition depends: whether—and to what extent—the satisfactory explanation of cognition requires classical computational mechanisms as opposed to nonclassical digital computation and whether we need to invoke processes that involve nondigital computation (Figure 1).

It may turn out that one computational theory is right about all of cognition, or it may be that different cognitive capacities are explained by different kinds of computation. To address these questions in the long run, the only effective way is to study

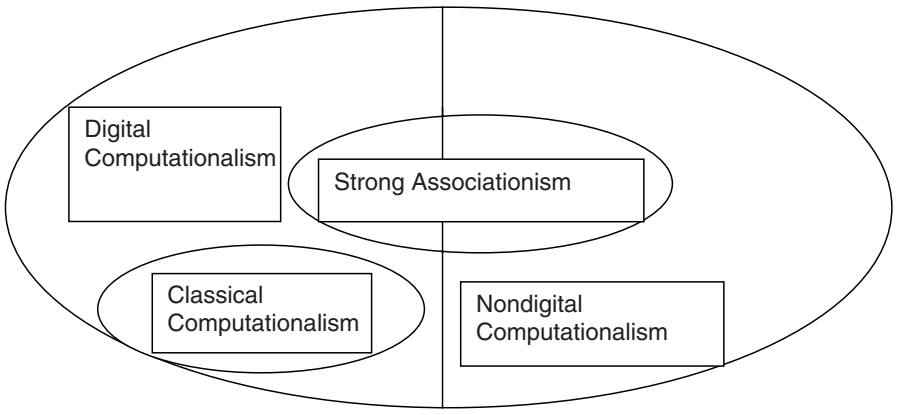


Figure 1 Some Prominent Forms of Computationalism and Their Relations

Source: Author.

nervous systems at all their levels of organization and find out how they exhibit cognitive capacities.

Gualtiero Piccinini

See also Artificial Intelligence; Cognitive Sciences; Machine Consciousness and Autonomous Agents; Mind–Body Relation; Models in Science; Social Neuroscience

Further Readings

Boden, M. A. (2006). *Mind as machine: A history of cognitive science.* New York, NY: Oxford University Press.

Ermentrout, G. B., & Terman, D. H. (2010). *Mathematical foundations of neuroscience.* New York, NY: Springer.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture. *Cognition, 28,* 3–71.

Householder, A. S., & Landahl, H. D. (1945). *Mathematical biophysics of the central nervous system.* Bloomington, IN: Principia.

McCulloch, W. S., & Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5,* 115–133.

Piccinini, G., & Scarantino, A. (2011). Information processing, computation, and cognition. *Journal of Biological Physics, 37*(1), 1–38.

Pylyshyn, Z. W. (1984). *Computation and cognition.* Cambridge: MIT Press.

Rumelhart, D. E., McClelland, J. M, & PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition.* Cambridge: MIT Press.

Schwartz, E. L. (1990). *Computational neuroscience.* Cambridge: MIT Press.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11,* 1–23.

Turing, A. M. (1965). On computable numbers, with an application to the Entscheidungs problem. In M. Davis (Ed.), *The undecidable* (pp. 115–153). New York, NY: Raven. (Original work published 1936–1937)

Coalition Logic

Coalition Logic (CL) is a modal logic for reasoning about cooperation, more specifically about *coalitional ability.* The main construct of CL is of the form $[C]\varphi$, where C is a group of agents (a coalition), meaning that coalition C can by some collective action bring about a state of the world where φ is true, no matter what other agents choose to do. In game-theoretic terms, $[C]\varphi$ means that C is *effective* for φ. Basic CL as well as epistemic, temporal, strategic, quantified, and other extensions of it can be used to formalize properties of social choice mechanisms in a natural way.

The Language of Coalition Logic

CL extends propositional logic with a family of modalities [C], one for each coalition C. Formally, the language of CL is parameterized by a finite set N of *agents* and a set of *atomic propositions*. *Formulae* are defined as θ follows. An atomic proposition $p \in \theta$ is a formula, and if φ and ψ are formulae, then $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$, $\varphi \rightarrow \psi$, $\varphi \leftrightarrow \psi$, and $[C]\varphi$ are formulae, where $C \subseteq N$. The meaning of the operators $\neg$, $\vee$, $\wedge$, $\rightarrow$, and $\leftrightarrow$ are as in propositional logic.

Coalition Logic and Social Choice

CL can be used to formally specify properties of mechanisms for social choice. As an example, consider the case that we are interested in a mechanism with the following properties. Two agents, a and b, should be able to collectively choose between one of two possible outcomes, p and q, but not both at the same time. However, none of the two agents should be able to choose any of the two outcomes individually. These properties can be formally specified in CL as follows: $a, bp \wedge a, bq \wedge \neg a, bp \wedge q \wedge \neg ap \wedge \neg bp \wedge \neg aq \wedge \neg bq$.

CL provides a formal language with a precise and well-defined meaning. In addition to making formal reasoning about social choice mechanisms possible in general, a formal logic also builds a bridge to the concept of *computation.* This connection, first, opens the door to computational tools developed in artificial intelligence and computer science to be used to automatically reason about the logical principles of a social mechanism, to specify and automatically verify properties of a social mechanism, and to automatically synthesize social mechanisms. Second, it is useful for understanding and implementing computational social choice mechanisms.

Semantics and Axioms

The meaning of CL formulae is defined in terms of neighborhood semantics. A *coalition model* is a tuple $M = \langle S,E,V \rangle$, where S is a set of *states*, $V{:}\theta \rightarrow 2S$ is a *valuation function* assigning atomic propositions to states, and for each state $s \in S$,

$E(s): \wp N \rightarrow \wp\wp S$ is an *effectivity function* assigning a set of sets of states to each coalition in state s. That $X \in EsC$ means that C can perform some collective action such that the outcome is ensured to be a state in X. Each effectivity function E(s) is required to be a *playable* effectivity function, that is, to have the following properties: (1) $\emptyset \notin E(s)(C)$, (2) $S \in E(s)(C)$, (3) $S \backslash X \notin Es\emptyset \Rightarrow X \in EsN$ (N-maximality), (4) $\forall X \subseteq X' : X \in EsC \Rightarrow X' \in EsC$ (outcome monotonicity), and (5) $\forall C1 \cap C2 = \emptyset : X1 \in EsC1$ and $X2 \in EsC2 \Rightarrow X1 \cap X2 \in EsC1 \cup C2$ (superadditivity).

These properties correspond exactly to a natural type of *effectivity* in strategic games. A *strategic game* is here a tuple $G = \langle N, \{\Sigma i : i \in N\}, o, S\rangle$, where N is the set of agents, S the set of states, Σi the nonempty set of strategies for i, and $o : \Pi i \in N \Sigma i \rightarrow S$ maps each strategy profile to an outcome state. The α-*effectivity function* EG of a strategic game G is defined as follows: $X \in EGC \Leftrightarrow \exists \sigma C \forall \sigma N \backslash Co(\sigma C, \sigma N \backslash C) \in X$. An effectivity function is playable if and only if it is the α-effectivity function of some strategic game and if it is assumed that the state space is finite. Without the finiteness assumption, an additional property is required for the correspondence to hold: (6) $\forall X \subseteq S : X \in EsN \Rightarrow \exists x \in X\ x \in E(s)(N)$ (crown).

Given a formula φ, coalition model M, and state s, we write $(M,s) \models \varphi$ means that φ is true in s, defined as follows: When $p \in \theta$, $(M,s) \models p$ iff $s \in V(p)$. The propositional connectives are interpreted as usual. Finally, $(M,s) \models [C]\varphi$ iff $\{s \in S : (M,s) \models \varphi\} \in EsC$.

The set of all *valid* formulae, true in every state in every coalition model, is completely axiomatized by the following axioms and inference rules.

All substitution instances of propositional tautologies
$\neg[C]\bot$ (where $\bot$ is a propositional contradiction)
$C\neg\bot$
$\neg[\emptyset]\neg\varphi \rightarrow [N]\varphi$
$[C](\varphi \wedge \psi) \rightarrow [C]\psi$
$([C1]\varphi 1 \wedge [C2]\varphi 2) \rightarrow [C1 \cup C2](\varphi 1 \wedge \varphi 2)$
From φ and $\varphi \rightarrow \psi$ derive ψ (*Modus ponens*)
From $\varphi \leftrightarrow \psi$ derive $C\varphi \leftrightarrow [C]\psi$ (Equivalence)

Extensions

Several extensions of the basic language have been proposed in order to, for example, be able to express more sophisticated properties of situations involving social choice. Alternating-Time Temporal Logic (ATL), developed independently from CL, extends CL syntactically with *temporal modalities* and semantically with *strategies*. *Epistemic* extensions of CL and ATL add epistemic modalities. Examples of formulae: $Ki\varphi \rightarrow iKj\varphi$ (i can communicate her knowledge of φ to j) and $CG\varphi \rightarrow G\psi$ (common knowledge in coalition G of φ is sufficient for G to ensure that ψ). Quantified Coalition Logic (QCL) adds quantification over coalitions. Consider a mechanism with the following properties. An electorate of n voters should be allowed to choose between two outcomes p and q, using majority voting. No nonmajority coalition should be allowed to choose an outcome, and any majority coalition should be allowed to choose any outcome. These properties can be succinctly described in QCL as follows: (majnp) $\wedge$ majnq $\wedge \neg\neg$majnp $\wedge \neg\neg$majnq. Here, majnp means that for *any* coalition C satisfying maj(n), that is, any majority coalition, Cp holds. $\neg$majnp means that *there exists* a coalition C satisfying $\neg$majn: that is, a nonmajority coalition such that Cp holds.

Thomas Ågotnes

See also Artificial Intelligence; Deontic Logic and Agency; Dialogical Logic; Epistemic Approaches to Democracy; Game-Theoretic Modeling; Modal Logic and Intentional Agency; Social Choice Theory

Further Readings

Ågotnes, T., van der Hoek, W., & Wooldridge, M. (2008). Quantified coalition logic. *Synthese, 165*(2), 269–294.

Goranko, V., Jamroga, W., & Turrini, P. (2011). Strategic games and truly playable effectivity functions. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems* (Vol 3., pp. 727–734). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Pauly, M. (2002). A modal logic for coalitional power in games. *Journal of Logic and Computation, 12*(1), 149.

Wooldridge, M., Ågotnes, T., Dunne, P. E., & van der Hoek, W. (2007). Logic for automated mechanism design: A progress report. In *Proceedings of the Twenty-Second Conference on Artificial Intelligence* (pp. 9–16). Vancouver, British Columbia, Canada: AAAI Press.

Cognitive Anthropology and Mental Architecture

Cognitive anthropology studies the collective structures, and their content, that make up culture. This entry introduces the novel field of cognitive anthropology; explains the importance of the notion of mental architecture in studying concept acquisition and concept extension, especially in social contexts involving essentially collective conceptual categories; highlights the importance of cognitive anthropology and cognitive sciences cross-fertilizing each other; and presents in detail the kind of groundbreaking approach to the collective cognitive structures underlying social interaction, particularly in different cultures, which this discipline has put forward.

"Mental architecture" can refer to several things, such as how the brain is structured, how the mind is structured, or how mental entities are structured, and how collective concepts (the properties of a community—distributed and varyingly shared within that community) are structured. Of this list, it is only the last item that is directly in the province of cognitive anthropology, though psychological theories about the structure of mind and mental entities are appealed to by some scholars, too. The idea of mental architecture is often associated with *modules,* or a modular approach wherein each entity is bounded, having its own structure.

Contemporary cognitive anthropology undermines certain traditional assumptions held by many students of semantics with regard to two topics. The first topic involves the explanation of certain universal conceptual categories—traditionally spoken of as "natural kinds": Here, the received view was that the acquisition of universal conceptual categories may be innate whether what is innate could be the knowledge of them itself or the attributes and contexts of knowledge by means of which we form these common concepts, such as, for example, "animate," or "living." The other traditional topic has to do with the nature of *collective cognitive* (including linguistic) *categories*. Is their structure one of categories bounded once and for all, or is it a "prototype-extension" structure, with much creative flexibility in patterns of extension—that is, permitting extension to the range of applicability of such concepts or, in other words, extending the denotation or referents of such concepts? Of these two assumptions undermined by contemporary developments, the latter is the one that involves a topic most basic to cognitive anthropology, namely, collective cognitive categories. Nevertheless, the former issue is also pertinent to our topic.

In this entry, after a discussion of the aforementioned undermined assumptions, three further issues important to any treatment of collective cognitive structures will be presented.

Natural Kinds (Undermining the First Traditional Assumption)

The *natural-kinds argument* states that there exist certain classes of natural, substantive categories—that is, concepts referring to physical substances, such as landscape features (e.g., mountains, rivers), living things, or plant genera (e.g., oak, pine), that we are innately programmed to recognize. Although it is natural to assume that there is such an innate disposition, since all people recognize something like these categories, the question concerns whether the knowledge of them itself (i.e., "this is what a plant genus looks like") is innate or whether, alternatively, we are born with only the ability to recognize attributes (e.g., mass, edges, continuity vs. discontinuity, apparently internally generated change over time, apparent volition, etc.) from which we further infer such categories.

The difference matters because some anthropologists and others have claimed that our recognition of a natural category, such as "oak," is intrinsically different from our recognition of a culturally constructed category, such as "table." What is thus crucial is that work within cognitive anthropology on semantic and pragmatic categories suggests that *no such difference exists*. Even nursery rhymes ("The fork ran away with the spoon") imply that when instances of inanimate, culturally created object categories (forks and spoons) are given the right attributes (such as volition and the ability to act) we have no trouble imagining those instances as animate. And complex machines (e.g., computers and cars) are sometimes felt to be animate even while we know, intellectually, that they are not (consider Hal, the rogue computer, in the movie *2001*).

Collective Cognitive Categories (Undermining the Second Traditional Assumption)

In the following discussion, "term" (cf. *morpheme* or *lexeme*) is used to refer to a *signifier* (a sound

concept) and "referents," to refer to various *signifieds* (concepts); both kinds of concepts may be "realized" in the world as experienced by language-using human beings. "Word" is used to refer to a *sign*—Ferdinand de Saussure's union of a *signifier* and a *signified* (a relationship often misunderstood in discussions of "structuralism"). Understanding the nature of the two Saussurean concepts and the relationship between them is important because the one-to-one connection between *signifier* and *signified* that characterizes "componential semantics" is broken in extensionist approaches, as discussed below.

Componential Categories

Classically, semantic categories were considered—by early cognitive anthropologists and those who strove for rigorous formal semantic definitions—to be bounded, "all and only" chunks of semantic space. This approach was embodied in the componential analysis of kin terminologies by early cognitive anthropologists: A term's set of referents is defined by the intersection of a set of semantic components. The only alternative to this formal view was seen as a kind of accidental concatenation of incidental links among the *signified*s tied to any particular *signifier,* captured by Ludwig Wittgenstein's idea of "family resemblances." The problem with the "unstructured incidental links" alternative is that our shared use of language is too complex and apparently regular; it has to depend on a system. However, the "all and only" categories alternative would leave us with woefully insufficient vocabularies. Since our vocabularies are individually learned but have to be learned in an identical way by enough of us to enable effective communication in social contexts, we do not have the experience or time that it would take to learn an adequately dedicated vocabulary for talking about all of the things we have occasion to communicate about; that is, in the normal course of events, we have occasion to speak of a great many more concepts than such a fixed vocabulary can capture. New speakers only learn words via their frequency of use in conversations, and it takes ample frequency to keep these words active in the language. Frequently spoken of innovations, whether technological (e.g., computers), social ("rave" parties), style related ("A-line"), or whatever, all lead to new signs in the language, while, in contrast, significantly less frequently used terms (such as "quill" for a pen, or "syke" or "stell" as kinds of watercourses) drop out of people's active vocabularies—even if, in modern literate languages, they are technically maintained in dictionaries. For example, we needed some way less awkward than the paraphrase "a new writing implement that works sort of like a pen" to talk of ball-point pens when they first began to emerge (before their own label became widespread and, thus, common).

Prototype Extension

Developing out of Floyd Lounsbury's studies of kinship semantics and O. Brent Berlin and Paul Kay's study of color terms, a third alternative approach to semantic categories has gradually emerged, the *prototype-extension* approach. In this approach, basic or default definitions apply to prototypic referents of terms, and it was the communicative usefulness of easy reference to these referents that generated the words in the first place. Prototypic referents are what terms are presumed to refer to in the absence of other information, and it is only to them that the "essential properties" of the words necessarily apply.

In this approach, *semantic extension* provides the solution for the problem of how we can talk of many more referents than we have specifically dedicated terms for. We are able to use familiar terms in systematic (and contextually sensitive) ways for referents that do not rate their own dedicated terms. *Denotative extension* is based on the similarity of the extended referent to the defining-form attributes of the prototype (joined with functional or connotative relevance); such extensions (e.g., extending the term *cousin* from the prototypic first cousin to more distant ones) are considered true referents of the term in question. *Connotative* extension is based on a direct similarity of function but where the extended referent lacks key attributes of the prototype (as when a senior family friend is called "uncle"). *Figurative* extension involves extending a term (with its key contrasts) from one domain to another in order to use the connotations of the term. Such extension can range from very conventional usage (say our use of "brother" variously for monks and for members of one's fraternity) to quite ad hoc metaphors (the "plumpness" of one's academic department as it awaits its looming fiscal "diet"). At the same time, such extensions must be sufficiently systematic and based on general understandings for hearers to be able to recognize (and

in most normal conversation, they must be able to *automatically* recognize) what the speaker intends them to communicate in context. Extension is not one that applies to a term in isolation but one that operates in the context of a term's focal contrasts (e.g., "brother" vs. "sister," "father," "son," etc., or "uncle" vs. "aunt," "father," "nephew," etc.). In this approach, conjunctive definitions remain important, but their focus of application is different from that in the componential approach.

In cognitive anthropology, *semantics* refers to the ways in which words are defined in structural linguistics and in the anthropology that grew out of it. These definitions are couched in terms of relations of contrast (or opposition), by means of which words are distinguished from one another, and inclusion, by means of which they are grouped into higher-level categories. Such information includes what distinguishes a "father" from an "uncle" or a "son"—as kinds of "relative" or "kin," an "oak" from a "pine"—as kinds of "trees," a "tree" from a "bush"—as kinds of "plant," and a "chair" from a "table"—as kinds of "furniture." But obviously our use of these terms depends on much other information, such as the facts that an "uncle" is a "father's" (or "mother's") brother, that mothers and fathers have special responsibilities for their children that uncles and aunts (in Western culture) do not have, and that an emotionally close older nonrelative, in America, can be spoken to or of as "uncle" but not "father." Similarly, we know that oak makes more solid furniture than does pine. This wider sort of presumptively shared knowledge about the world that informs our use and understanding of words can be spoken of as *pragmatics*. Pragmatics thus defined is *not limited to language* (as is semantics) but extends out to the similar kinds of presumptively shared knowledge we have of how to treat an uncle or a friend, or a person we are standing next to in a supermarket line. This wider universe of knowledge that both our speech and our social interaction (whether cooperation, competition, or ignoring) depend on can be seen as the *cognitive* version of what anthropologists speak of as "culture."

Three Further Issues

Kinds of Collective Cognitive Structures

From the perspective of cognitive anthropology, both semantics and pragmatics represent kinds of *collective cognitive structures,* emphasis being placed on all three terms, jointly and individually, at the same time: *collective, cognitive,* and *structure*—that is, mental architecture.

Semantics can be seen as a single system paralleling the more exclusively linguistic systems of phonology, syntax, and perhaps the lexicon. But pragmatics can be broken into a number of distinct subsystems (or modules); these include inter alia Edward Hall's *proxemics* (the ways people in different cultures space themselves out in different social situations or how space is perceived in different cultural contexts), cultural-values systems, culturally shaped modes of emotional response (Gregory Bateson's *ethos*), "cultural modes of thought" (conceptual *bauplan*s—basic ways of imagining and ordering conceptual relations in a culture (cf. Bateson's *eidos*), "cultural conceptual systems" (formally organized conceptual systems such as folk taxonomies or kinship terminologies), and "cultural models of action" (models for how to behave in a given situation, e.g., going on a date or building a house—often spoken of simply as "cultural models," associated with the work of Naomi Quinn and Claudia Strauss, among others). Early cognitive anthropology had a significant impact on the initial development of the new discipline of cognitive sciences, and more recently, insights from cognitive sciences have helped shape developments in cognitive anthropology.

Collective Knowledge Systems and Individual Knowledge

The relationship between collective knowledge systems and individual knowledge is a tricky one. In any direct sense, all there is is individual knowledge, since there exists no collective mind or collective consciousness. But at the same time, we all constantly act as if there exist *collectively* held cognitive systems—such as the proper grammar of a language, the proper pronunciation and meaning of words, the correct way to show respect for an elder, what counts (socially and culturally, vs. legally) as justifiable homicide, and so forth. We presume these collective systems in our communication and interaction. What makes these putative collective systems effectively real is our constant reference to (or dependence on) them, joined with the fact that we know how they differ from our personal systems (e.g., "my accent" or "I prefer to say it this way") and the fact that we agree more with each other

about what are generally considered "correct" or "normal" usage, views, values, and so on (even if we ourselves do not necessarily adhere to these) than we do regarding our personal actions or preferences.

As individuals learning a cultural or linguistic system, we keep trying to construct a representation of it in our minds that allows us to generate acceptable performance and judgments, enabling us to make sense of the messages and behavior that we experience from others. Mostly, this learning process and the knowledge it produces are subconscious, but occasionally bits of the learning or the knowledge on which we are momentarily focused will rise to consciousness. We thus each have in our minds our *individual* knowledge (i.e., individual representations) of each of the various *collective* representations that make up our language and culture—including dialects, subcultures, and so forth, and the system variants that go with each social entity within which we participate or with whose members we interact.

Our separate individual representations of these communal cognitive systems are kept consistent with each other because our social interactions depend on them—but only to the degree to (and in the situations in) which that interaction is actually relevant. Thus, we all pretty much agree on what unicorns look like but don't seem much to agree on the nature of their feces; and more prosaically, English speakers agree on what a cousin is, but that agreement falls apart when we get into kinds of cousins (such as "first cousin, once removed" vs. "second cousin"). Note that we are talking about basic cultural *knowledge,* not about matters of taste or opinion—which vary even more. The details of how such systems are learned, organized, kept straight, and stored are topics of ongoing study.

Cognitive Structures and Social Units

For structures to be collective, they have to inhere in some kind of social units. At the same time, it is the shared systems of language and culture that enable membership in social units to be recognized and to function. In this sense, culture (including language) and society can be said to be mutually constitutive. However, collective cognitive structures pertain not just to society as a whole but to the full gamut of (and variety of hierarchies of) social units, whether formal or informal, face-to-face or indirect, continuing or ad hoc; and so they delimit a matching variety of pragmatic subcultural cognitive structures. The anthropological study and modeling of the systematic relationships among such subcultural cognitive structures is still in its infancy.

David B. Kronenfeld

See also Cognitive Sciences; Collective Agents; Concepts; Distributed Cognition and Extended-Mind Theory; Group Identity; Kinds: Natural Kinds Versus Human Kinds; Language and Society; Semantics and Pragmatics; Social Anthropology; Sociolinguistics; Structuralism and Poststructuralism

Further Readings

Bateson, G. (1972). *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology.* Chicago, IL: University of Chicago Press.

Bennardo, G. (2009). *Language, space and social relationships: A foundational cultural model in Polynesia.* Cambridge, England: Cambridge University Press.

Berlin, O. B., & Kay, P. (1969). *Basic color terms.* Berkeley: University of California Press.

D'Andrade, R. G. (1995). *The development of cognitive anthropology.* Cambridge, England: Cambridge University Press.

D'Andrade, R. G., & Strauss, C. (Eds.). (1992). *Human motives and cultural models.* Cambridge, England: Cambridge University Press.

Hall, E. T. (1966). *The hidden dimension.* Garden City, NY: Doubleday.

Kronenfeld, D. B. (1996). *Plastic glasses and church fathers: Semantic extension from the ethnoscience tradition* (Oxford Studies in Anthropological Linguistics). New York, NY: Oxford University Press.

Kronenfeld, D. B. (2008a). Cultural models. *Intercultural Pragmatics, 5,* 67–74.

Kronenfeld, D. B. (2008b). *Culture, society, and cognition: Collective goals, values, action, and knowledge* (Mouton Series in Pragmatics 3). Berlin, Germany: Mouton de Gruyter.

Kronenfeld, D. B., Bennardo, G., De Munck, V. C., & Fischer, M. D. (Eds.). (2011). *A companion to cognitive anthropology.* Malden, MA: Wiley-Blackwell.

Lounsbury, F. G. (1964). A formal account of the Crow- and Omaha-type kinship terminologies. In W. H. Goodenough (Ed.), *Explorations in cultural anthropology* (pp. 351–393). New York, NY: McGraw-Hill.

Lounsbury, F. G. (1969). Language and culture. In S. Hook (Ed.), *Language and philosophy* (pp. 3–29). New York, NY: New York University Press.

Saussure, F. de (1959). *Course in general linguistic* (C. Bally & A. Sechehaye [with A. Reidlinger], Eds.; W. Baskin, Trans.). New York, NY: Philosophic Library. (Original work published 1916)

Strauss, C., & Quinn, N. (1997). *A cognitive theory of cultural meaning*. Cambridge, England: Cambridge University Press.

Cognitive Archaeology

Cognitive archaeology, or the archaeology of mind, has been broadly defined by Colin Renfrew as the study of past *ways of thinking* through the material remains of the past. What does this mean? Why ask an archaeologist about the mind? Above all, what is it, specifically, that archaeology can bring to the study of mind and human cognitive evolution?

Two major points may be made by way of answer: First, as our stated working definition makes clear, the archaeology of mind is not concerned with *what* people were thinking in the past but, instead, with *how* they were doing their thinking. Cognitive archaeology is primarily interested, in other words, to learn about the different situated processes, rather than disembodied contents, of human thought as they emerge and take shape in the different cultural trajectories of our species. This is important because as archaeology may well testify, significant episodes of the story of mind appear relatively recently in the archaeological record and can certainly be seen as the emergent products of various cultural trajectories rather than innate biological capacities. Indeed, the image of the human mind that emerges from the perspective of cognitive archaeology so far resembles that of a hybrid bio-cultural entity constituted in a double way: It is partly the product of biological evolution and partly an artifact of our own making.

How did we humans get to be so different then? This brings us to our second major point, which can be found by looking at three key topics that characterize, more than anything else, the archaeology of mind as an *epistemic culture:* time, change, and material culture. Starting with the early Paleolithic stone tools some 2.6 million years ago and moving to the more recent and ever more refined forms of artifacts and technologies, cognitive archaeology sets out not only to understand the workings of the minds that created them but also to explore the possible role that these objects might have played in extending our minds and redesigning ourselves. This coevolutionary process of deep enculturation and material engagement, which continues into the present, occupies the principal focus for cognitive archaeology. The recently developed field of *neuroarchaeology,* in particular, is primarily preoccupied with understanding the long-term developmental mechanisms by which the bidirectional, mutual constitution of brain and culture occurs. To this end, a basic commitment to some broadly defined, embodied and interactionist view of mind might protect us from sliding into neural determinism and thus from mistaking the properties of the sociocultural system for the properties of the individual or the brain.

Meanwhile, cognitive archaeology's special preoccupation with the world of things and long-term change signifies not only what differentiates it from the other disciplines in the field of contemporary cognitive sciences but also the point of intersection between these disciplines, to which cognitive archaeology may have some interesting contributions to make. For instance, one important contribution can be seen in the way cognitive archaeology has made clear that the science of mind and the science of material culture are two sides of the same coin. It is now understood that the way we think is not only changed but also partly constituted through the social experience of action and material engagement activities. New forms of engagement with the material world bring about new possibilities and constraints for the mind. Another, and closely related, contribution of cognitive archaeology can be seen in the recognition that the hallmark of human cognitive evolution may be based not on the ever-increasing sophistication or specialization of a modular mind but on an ever-increasing representational flexibility that allows for environmentally and culturally derived plastic changes in the structure and functional architecture of the human brain.

Naturally, the task of trying to reconstruct the prehistory of mind has not been an easy one. Establishing testable, empirical, but also culturally sensitive and philosophically informed, links between brain structure, cognitive function, and

archaeologically observable behaviors remains a central challenge. But since the emergence of cognitive archaeology in the early 1980s, there have been some radical advancements as well as changes in perspective in the way we approach, interpret, and understand the ancient mind *through the material record*. Especially in the past decade or so, the archaeology of mind has apparently come of age with a series of remarkable theoretical advances and a growing number of new empirical findings. The initial skepticism about "paleopsychological" investigations gave way to a systematic research endeavor that fuses different schools of thought and research strands together in a highly interdisciplinary and rapidly growing research field.

New, important steps have been made toward an understanding of the basic principles of brain evolution and of human brain–culture coevolution. As a result, we can now make better sense of the evolutionary changes (structural/functional) that occurred, the ways in which human brains differ, as well as the manner in which these differences evolved. In this context, new methodological advances in the field of paleoneurology enable a more nuanced interpretation of the observed variation in endocranial morphology and a better understanding of the patterns of growth (changes in size) and development (changes in shape) of the brain as reflected in the available record of fossil endocasts. Moreover, a new hypothesis on the role of enhanced working memory in the emergence of modern human cognition has been proposed.

Last but not least, important theoretical developments in phenomenology and cognitive science assist us in rethinking our conventional ideas about the boundaries between cognition and material culture, and the embodied character of human experience. Archaeologists have learned a great deal about the social and cultural basis of the human mind while working with anthropologists, philosophers, and social and cognitive scientists. Conversely, cognitive archaeology, probably more than any other field of cognitive research, has the ability to develop a systematic understanding of the relationship between cognition and material culture from its own unique perspective. The temporal depth and historical diversity of the archaeological object, coupled with this systematic examination of the interaction between mind and matter in the course of human cognitive evolution, may well furnish some key insights into the nature of human condition: who we are and how we got here.

Lambros Malafouris

See also Actor-Network Theory; Cognitive Anthropology and Mental Architecture; Cognitive Sciences; Cultural Evolution; Embodied Cognition; Evolutionary Psychology; Modularity of the Mind; Situated Cognition; Social Neuroscience

Further Readings

Coolidge, F., & Wynn, T. (2009). *The rise of Homo sapiens: The evolution of modern thinking*. Oxford, England: Wiley-Blackwell.

de Beaune, S., Coolidge, F., & Wynn, T. (Eds.). (2009). *Cognitive archaeology and human evolution*. Cambridge, England: Cambridge University Press.

Donald, M. (1991). *Origins of the modern mind*. Cambridge, MA: Harvard University Press.

Gamble, C. (2007). *Origins and revolutions: Human identity in earliest prehistory*. Cambridge, England: Cambridge University Press.

Knappett, C. (2005). *Thinking through material culture: An interdisciplinary perspective*. Philadelphia: University of Pennsylvania Press.

Knappett, C., & Malafouris, L. (Eds.). (2008). *Material agency: Towards a non-anthropocentric approach*. New York, NY: Springer.

Malafouris, L. (2004). The cognitive basis of material engagement: Where brain, body and culture conflate. In E. DeMarrais, C. Gosden, & C. Renfrew (Eds.), *Rethinking materiality: The engagement of mind with the material world* (pp. 53–62). Cambridge, England: McDonald Institute for Archaeological Research.

Malafouris, L. (2008). Beads for a plastic mind: The "blind man's stick"(BMS) hypothesis and the active nature of material culture. *Cambridge Archaeological Journal, 18,* 401–414.

Malafouris, L. (2009). "Neuroarchaeology": Exploring the links between neural and cultural plasticity. *Progress in Brain Research, 178,* 251–259.

Malafouris, L., & Renfrew, C. (2010). The cognitive life of things: Archaeology, material engagement and the extended mind. In L. Malafouris & C. Renfrew (Eds.), *The cognitive life of things: Recasting the boundaries of the mind* (pp. 1–12). Cambridge, England: McDonald Institute for Archaeological Research.

Mithen, S. (1996). *The prehistory of mind*. London, England: Thames & Hudson.

Noble, W., & Davidson, I. (1996). *Human evolution, language and mind: A psychological and archaeological inquiry.* Cambridge, England: Cambridge University Press.

Renfrew, C. (2006). Becoming human: The archaeological challenge. *Proceedings of the British Academy, 139,* 217–238.

Renfrew, C. (2007). *Prehistory: The making of the human mind.* London, England: Weidenfeld & Nicolson.

Renfrew, C., Frith, C., & Malafouris, L. (2008). Introduction. The sapient mind: Archaeology meets neuroscience. *Philosophical Transactions of the Royal Society of London Series B, 363,* 1935–1938.

Renfrew, C., Frith, C., & Malafouris, L. (Eds.). (2009). *The sapient mind: Archaeology meets neuroscience.* Oxford, England: Oxford University Press.

Renfrew, C., & Scarre, C. (Eds.). (1998). *Cognition and material culture: The archaeology of symbolic storage* (McDonald Institute Monographs). Cambridge, England: McDonald Institute for Archaeological Research.

Stout, D., Toth, N., Schick, K., & Chaminade, T. (2008). Neural correlates of Early Stone Age toolmaking: Technology, language and cognition in human evolution. *Philosophical Transactions of the Royal Society of London Series B, 363,* 1939–1949.

Wynn, T., & Coolidge, F. L. (2008). A stone-age meeting of minds. *American Scientist, 96*(1), 44–51.

Cognitive Phenomenology

The term *cognitive phenomenology* refers to the idea that the phenomenal quality of experience is not limited to the sensory or feeling states but that there is, in Thomas Nagel's phrase, "something it is like" to engage in higher-order cognitive acts like thinking, judging, or evaluating. The claim that there is a phenomenal or experiential aspect to thinking is controversial, with some philosophers arguing against the idea (e.g., David Braddon-Mitchell and Frank Jackson, Peter Carruthers, Michael Tye) and others arguing in favor of it (e.g., Terence Horgan and John Tienson, Charles Siewert, Galen Strawson, or Dan Zahavi). A related but distinguishable issue is whether phenomenal consciousness is conceptual or nonconceptual or, more precisely, whether the conceptual aspects of thinking contribute something to phenomenal experience even if the phenomenality in question is sensory.

Those who support the idea that thinking does involve a phenomenal or experiential aspect claim that just as there is something it is like to taste a lemon, there is something it is like to think about a lemon and that these two phenomenal states are different. There is a difference between what it is like to actually taste a lemon and what it is like to remember what it is like to taste a lemon, just as there is a difference between what it feels like to throw a lemon and what it feels like to calculate how far I might be able to throw it.

Intentional states have a phenomenal character, and this phenomenal character is precisely the "what it is like" of experiencing a specific propositional attitude vis-à-vis a specific intentional content. Change either the attitude type (believing, desiring, wondering, hoping, etc.) or the particular intentional content, and the phenomenal character thereby changes too.

Likewise, Alvin Goldman suggests that when we think about something, we are implicitly (nonreflectively) aware of our thinking. This seems to imply that thinking has a phenomenal feel to it. This may mean that there is a specific qualitative feel to having a specific thought, or it may mean that thought processes, like associating or inferring or speculating, may have different feels. One might claim, however, that there is *nothing* that it is like to think, except that one is aware that one is thinking. Nor is it clear how specific the awareness or feeling is. Solving a math problem may feel different from planning a road trip, and engaging in either of these cognitive tasks may feel different from evaluating the motives of your best friend. But does solving a simple arithmetical problem feel different from doing calculus? Does planning a road trip to Memphis feel different from planning a holiday trip to Crete? Does evaluating someone's motives feel different from evaluating their physical health? On the one hand, even if one cannot pick out a phenomenal difference in every thought (or attitude type), the lack of phenomenal difference does not necessarily mean a lack of phenomenality. If we say that Thought A feels no different from Thought B, this doesn't mean that they have no feel at all to them. On the other hand, if one can pick out a difference in each of these thoughts, then it seems possible that the phenomenal experience is linked to the intentionality or the content of thinking.

One could take this idea further and claim that the difference between doing arithmetic and doing calculus is not about the doing but about the complexity

of the problem that's being considered, or that what it feels like to think about something depends entirely on the something that is being thought about. In this regard, there are three possibilities: (1) the differences under consideration are not differences in the phenomenal aspects of thinking but rather simply differences in intentional content, and the phenomenal aspects are completely distinguishable from the intentional aspects; (2) the differences under consideration are differences in the phenomenal aspects of consciousness but are linked to intentional differences because the phenomenal and intentional aspects are related or even inseparable; and (3) the differences are in some way purely phenomenal differences that don't depend on intentional content.

Those who argue against the idea that thinking has a phenomenal feel suggest that the idea confuses thinking with the possible inner speech or visual imagery that might accompany thinking. Inner speech or visual imagery may be phenomenally conscious, but the thinking itself intrinsically (i.e., nonrelationally) has no phenomenal quality. But one might object that someone who rehearses a bit of inner speech, for example, "La porte est rouge," but who isn't thinking that the door is red, because she doesn't know what these words mean, would have a different phenomenal experience from someone who understands the words. In that case, the thinking itself seems to make a difference. The question, as Carruthers and Veillet (2011) have recently put it, is whether, with regard to some phenomenal experience, a concurrent thought "makes a constitutive, as opposed to a causal, contribution to the phenomenal properties" (p. 3). On their view, no one would deny that thinking about something may play a causal role in generating phenomenal experience. The central question, however, is whether the thinking is constitutive of the phenomenal experience, that is, whether thinking is instrinsically phenomenal or not.

The debate about cognitive phenomenology is ongoing. Both sides have appealed to introspection, both have appealed to inference to the best explanation, different sides have contested whether phenomenal aspects of cognition facilitate self-knowledge (knowing what mental state one is in), and disagreements abound about the importance of this issue to questions about representational content. Everyone does agree, however, that the purported fact that making arguments on one side feels more convincing than making arguments on the other side, which apparently involves no feeling at all, is not decisive.

Shaun Gallagher

See also Consciousness; Embodied Cognition; Intentionality; Introspection (Philosophical Psychology); Self and the Social Sciences; Self-Knowledge; Social Cognition

Further Readings

Bayne, T., & Montague, M. (2011). *Cognitive phenomenology.* Oxford, England: Oxford University Press.

Braddon-Mitchell, D., & Jackson, F. (2007). *Philosophy of mind and cognition* (2nd ed.). Oxford, England: Blackwell.

Carruthers, P. (2005). *Consciousness: Essays from a higher-order perspective.* Oxford, England: Oxford University Press.

Carruthers, P., & Veillet, B. (2011). The case against cognitive phenomenology. In T. Bayne & M. Montague (Eds.), *Cognitive phenomenology* (pp. 35–56). Oxford, England: Oxford University Press.

Gallagher, S. (2005). *How the body shapes the mind.* Oxford, England: Oxford University Press.

Goldman, A. (1970). *A theory of human action.* New York, NY: Prentice Hall.

Horgan, T., & Kriegel, U. (2008). Phenomenal intentionality meets the extended mind. *The Monist, 91,* 353–380.

Horgan, T., & Tienson, J. (2002). The intentionality of phenomenology and the phenomenology of intentionality. In D. J. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 520–533). Oxford, England: Oxford University Press.

Kripke, S. (1982). *Wittgenstein on rules and private language.* Oxford, England: Blackwell.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83,* 435–450.

Siewert, C. P. (1998). *The significance of consciousness.* Princeton, NJ: Princeton University Press.

Strawson, G. (1994). *Mental reality.* Cambridge: MIT Press.

Tye, M. (1995). *Ten problems of consciousness.* Cambridge: MIT Press.

Tye, M. (2000). *Consciousness, color and content.* Cambridge: MIT Press.

Tye, M., & Wright, B. (2011). Is there a phenomenology of thought? In T. Bayne & M. Montague (Eds.), *Cognitive phenomenology* (pp. 326–344). Oxford, England: Oxford University Press.

Zahavi, D. (2006). *Subjectivity and selfhood: Investigating the first-person perspective.* Cambridge: MIT Press.

Cognitive Sciences

Cognitive science is the interdisciplinary attempt to understand mind and intelligence by combining insights from psychology, neuroscience, linguistics, artificial intelligence, and anthropology.

What are the relations among the cognitive sciences, such as psychology and neuroscience, and the social sciences, such as economics and sociology? This broad question raises many subordinate issues, such as the following: What distinguishes the cognitive sciences from the social sciences? Are theories and explanations in the cognitive sciences different from those in the social sciences? Can the cognitive sciences be reduced to the social sciences, or are the social sciences methodologically independent from or even conceptually prior to the cognitive sciences? What is the relation between social entities such as companies and nations and the individual people in them?

This entry outlines some of the main answers that have been given to these questions, thus tracing the intricate relationship between cognitive and social sciences.

Origins

The term *cognitive science* was coined in the 1970s to advocate a common set of concerns and ideas that form the intersection rather than just the union of all the fields that might be called cognitive sciences. There is no analogous interdisciplinary field called *social science* but rather a host of largely independent disciplines, such as sociology, economics, political science, anthropology, and psychology. Several fields are often placed among *both* the cognitive and the social sciences, including psychology, anthropology, and linguistics. This double placement is natural for fields that are concerned with the operations of individual minds *and* with the operations of minds in social contexts, as occurs in subfields such as social psychology, cognitive anthropology, and sociolinguistics.

Cognitive Sciences and Social Sciences: Relationship

Despite this overlap, there is an obvious difference in primary focus between the cognitive and social sciences, with the former concerned with the mental processes of individuals and the latter concerned with the social processes (mostly) of groups, such as companies and nations. Hence, we can use the term *mind–group problem* for the set of issues about the relation between the cognitive and social sciences. Progress in both the cognitive and the social sciences requires dealing with these issues, as is evident from recent interdisciplinary developments. For example, cultural psychology has been challenging the presuppositions of individualist cognitive psychology, and behavioral and neural economics have begun to provide alternatives to traditional approaches to economic analysis.

Dealing with the mind–group problem must begin with an understanding of the nature of the theories that are used to explain cognitive and social processes. Little is to be gained by viewing theories as sets of formalized universal statements, as the logical positivists advocated, because theories in both the cognitive and the social sciences rarely have this kind of structure. Instead, there is growing appreciation that cognitive and social theories can be better understood as descriptions of mechanisms, which are systems of parts whose interconnections produce regular changes. From this perspective, explanations of both cognitive and social phenomena consist of describing the mechanisms—parts and interactions—that causally produce the changes that require explanation. The relevant mechanisms are not just the simple push–pull interactions of Newtonian physics but can involve all the complexity of biological systems, including feedback relations and chaotic unpredictability. Then the mind–group problem takes on this form: What are the relations among the mental mechanisms relevant to explaining individual behavior and the social mechanisms relevant to explaining social behavior? This formulation is obviously controversial for practitioners of the cognitive and social sciences, who claim that explanations of human behavior cannot be mechanistic because they need to rely on introspection of self or empathic interpretation of others.

Three Views of the Relation Between Cognitive and Social Sciences

There are several alternative views of the relation between cognitive and social mechanisms: cognitive reduction, social construction, autonomy, and interaction. Reductionism is the view that social changes

can be explained in terms of psychological mechanisms, which in turn can be explained by neural mechanisms. Social constructionism, in contrast, makes social explanations fundamental on the grounds that all knowledge is socially constructed. Autonomy is the view that the cognitive and social sciences do and should operate independently from each other, so that neither psychology nor sociology, for example, needs to pay attention to ideas from the other field. The interactionist view may be seen as preferable, namely, that there are many important connections between the cognitive and social sciences that can be understood in terms of interacting mechanisms.

In order to adjudicate among these views, we need multiple examples of the kinds of mechanisms that currently appear to have the most explanatory potential at both individual and group levels. Modern approaches to cognitive science arose in the 1950s because ideas about computing provided an alternative to behaviorist explanations, whose inadequacy was increasingly becoming evident. Behaviorists in psychology and linguistics thought that explanation of the performance of humans and other animals should only concern how behavioral responses result from environmental stimuli that produce learning by reinforcement. Psychologists such as George Miller drew on new ideas about computer programs to argue that behavior could be better explained by supposing that animals from humans to rats employ mental representations and algorithmic processes that operate on the representations. This approach is clearly mechanistic: The parts are representations, the interactions are computational operations, and the behavioral changes result from applying the operations to the representations. For example, one important kind of explanation in cognitive science proposes that the most important mental representations are IF-THEN rules that can be used to solve problems by means of algorithms for selecting and applying rules in particular situations. A navigational problem can be solved by reasoning that operates with rules such as "IF you want to get to Australia, THEN arrange a flight."

Advances in Cognitive Science

Connectionism or Parallel Distributed Processing

Many other kinds of representations have also been proposed as important to explaining human cognition, including concepts, images, and analogies. These representations also function as parts of systems for processing information via computational interactions. In the 1980s, a new approach to cognitive explanations became influential, called *connectionism* or *parallel distributed processing*. Inspired in part by neural processing in the brain, connectionists view thinking as resulting from simple neuron-like processers connected by excitatory and inhibitory links. Computational models using neural networks, rules, concepts, images, and analogies have been used to simulate and thereby explain a wide range of kinds of thinking, including inference, problem solving, learning, and language use.

Contemporary Advances: Cognitive Neuroscience

In the 21st century, by far the most striking trend in cognitive science is the increasing prominence of *cognitive neuroscience,* which has both experimental and theoretical sides. The experimental side has exploded because of the availability of brain-scanning technologies such as functional magnetic resonance imaging. These technologies make it possible to examine what is happening in the brain while people perform complex tasks such as solving problems. Brain-scanning experiments produce vast amounts of data that need to be explained using new theories about how brains process information. The field of theoretical neuroscience develops computational models that simulate how interactions of large numbers of neurons can produce mental changes. These models are more biologically accurate than the earlier connectionist models, through employment of neural networks that are closer to those in the brain with respect to firing and connection properties. Explanations of thinking employed by cognitive neuroscience are obviously mechanistic: The parts are neurons, the interactions are the electrical and chemical ways in which neurons influence each other, and the resulting changes correlate well with the mental changes revealed in behavior.

Brain–Mind and Brain–Mind–Group

These advances in cognitive neuroscience make all the more pressing the traditional philosophical issue of the relation between body and mind, which becomes the *brain–mind problem*. This problem concerns the relation between neural mechanisms and mental ones and hence is similar to the mind–group problem concerning mental mechanisms and

social ones. Connecting these together, we get the *brain–mind–group* problem, whose solution would provide answers to a host of questions concerning the cognitive and social sciences. Before approaching this extended problem, we need a brief characterization of the kinds of social mechanisms that need to be connected with mental and neural ones.

Social Mechanisms

Social scientists are much less accustomed than biologists or cognitive scientists to describe explanations in terms of mechanisms, but it is easy to view social processes as consisting of interacting parts. At the first level, the parts are persons, and the interactions are all the forms of communication that occur between them. These include not just verbal conversations but also electronic messages and the full range of nonverbal communication, such as the transfer of emotional information by bodily signals. The social interactions that result in social changes can take many forms, such as power relations in which some individuals use verbal or nonverbal communication to induce emotional reactions in others that make them act in ways that the powerful want. For example, a political leader can wield power over a population in many ways, such as making them fear a foreign power or convincing them that the leader is the solution to their economic problems. Hence, power is a social mechanism that depends on cognitive mechanisms by which the powerful induce emotional reactions in the weak. Higher levels of social mechanisms concern group–group interactions, for example, in international relations where multiple countries cooperate and compete.

Neural, Psychological, and Social Mechanisms

If that understanding of social mechanisms is correct, we can approach the brain–mind–group problem by considering the relations among neural, psychological, and social mechanisms. There are two main reasons for rejecting the reductionist views that social phenomena are fully explainable by psychological mechanisms and that psychological phenomena are fully explainable by neural mechanisms. First, the phenomena to be explained are so complicated that it is difficult to see how a full account of them could be given just in terms of lower-level mechanisms. Consider, for example, social movements such as feminism and environmentalism. These involve the interactions of large numbers of people formed into identifiable groups that help determine who interacts with whom. There is no obvious way to identify social entities such as groups with psychological or neural entities. Second, causal explanation does not always operate in the reductionist direction from lower mechanisms to higher phenomena but sometimes can proceed from higher mechanisms to lower phenomena. For example, it might be a part of the explanation of why there is neural firing in the dopamine pathways of a fervent environmentalist to say that he or she is happy to be interacting with other members of an activist organization. In this case, the social causes the neural.

Recognition of social causation is not the imperialistic position found in some postmodernist social scientists that *all* knowledge is socially constructed, which implies that psychology and the other cognitive sciences can be ignored. Social processes undoubtedly contribute to all scientific developments, but so do psychological processes such as problem solving, as well as the neural processes that enable human brains to think. Experimentation in psychology and neuroscience, not to mention physics and chemistry, is sufficiently robust for the claim that scientific knowledge is *only* socially reconstructed to be hugely implausible.

Cognitive and Social Sciences: Interactions

Perhaps, then, we should conclude that the cognitive and social sciences are simply independent of each other and can pursue their own agendas. In the early decades of cognitive science, it was common to suppose that psychology was largely autonomous from neuroscience in that it dealt abstractly with information processes in common with computers, concerned with software functions rather than the underlying hardware. Advances in brain-scanning technologies, however, have utterly undermined this view, and since the 1990s, cognitive psychology has increasingly been integrated with neuroscience. Similarly, there are growing signs of recognition of the relevance of the cognitive sciences to the social sciences, for example, in the subfields of cognitive sociology and neuroeconomics. The social sciences are not so successful in their explanations and predictions that they can afford to ignore what psychology and neuroscience have to say about how people behave in social contexts. For example, the financial

collapse of 2008 was inexplicable in terms of the economic views of idealized individuals assumed by rational choice theories, but it begins to make sense from the perspective of psychological theories of human cognitive and emotional limitations.

In the other direction, the cognitive sciences do not hesitate to draw on the social sciences for many kinds of explanatory factors that contribute to explanations of how people behave in interaction with others. There is increasing recognition that social processes are an important part of intelligence, even in computer science, where the subfield of multi-agent systems looks at how artificial intelligence can be distributed across multiple interacting machines. Hence, the autonomy view is implausible as an account of the primary relation between the cognitive and the social sciences.

The weaknesses in the reductionist, social constructionist, and autonomy views show the need for the development of a richer account of how the cognitive and social sciences can fruitfully interact. Consider the important phenomena that involve ideological change, for example, the rise of new social movements such as communism and fascism or, more recently, the Tea Party and Occupy Wall Street interest groups in the United States. Such ideological developments are social phenomena observable in rallies and demonstrations, but they are also cognitive phenomena rooted in the beliefs and attitudes of the participants in the movements.

Cognitive and Social Mechanisms: Interrelation

What cognitive mechanisms are responsible for the adoption and maintenance of ideologies by individuals? Ideologies are appealing as the result of processes of emotional coherence, in which people form their beliefs and attitudes in response to their goals as well as the available evidence. The main psychological mechanisms include motivated inference, in which people marshal evidence in ways that suit their goals, and fear-driven inference, in which people arrive at beliefs that scare them because anxiety causes them to focus on the limited evidence for those beliefs.

These cognitive mechanisms are complemented by social mechanisms that transmit both cognitive and emotional forms of information. Here are some of the relevant mechanisms that bridge the cognitive and the social:

1. *Verbal communication,* in which people make statements and perform other kinds of speech acts
2. *Neural mirroring,* in which observation of the experience of another person can produce the same neural activity that would be produced by having the experience oneself
3. *Emotional contagion by mimicry,* in which people mimic the facial expressions of those with whom they interact, inclining them to acquire similar emotional reactions because emotions are in part responses to bodily changes
4. *Attachment-based learning,* in which emotional attitudes are acquired from people, such as parents, to whom a person is emotionally attached
5. *Empathy,* in which people acquire an emotional response from others by imagining themselves in the others' situations and experiencing emotions similar to theirs
6. *Altruism and sympathy,* in which people can acquire emotional responses directed toward the well-being of others
7. *Social cuing,* in which people's facial expressions, such as anger, can cue negative emotions, such as guilt, in their targets
8. *Power manipulations,* in which one person gains power over others by offering them something they desire or by offering to protect them from something that they fear

All of these social communication mechanisms depend on underlying psychological and neural mechanisms but do not reduce to them, because their description requires reference to the social processes of different people interacting with each other. Hence, explaining the spread of an ideology through a population requires attention not only to the cognitive/emotional mechanisms operating in individual brains but also to important social events, such as conversations, meetings, rallies, demonstrations, and occupations. These events provide the occasions by which social mechanisms ranging from verbal communication to power manipulations can affect the psychological and neural mechanisms operating within individuals.

The spread of ideologies in groups exemplifies the principle that the actions of groups result from the actions of individuals who think of themselves as

members of groups. On this principle, the explanation of group behaviors is

(a) not individualist, because groups are not reduced to mental representations;
(b) not holistic, because the psychology of individuals is considered;
(c) emergentist, because the actions of groups are different from the properties of individuals and depend on complex interactions between the individuals, including their representations of the group and each other.

Emergence

The concept of emergence is a hotly contested one in the philosophy of science, but a reasonable and nonmystical version is available: Emergent properties belong to the wholes, do not belong to any of the parts, and are not aggregates of properties of the parts. A social group, such as a political party, can have emergent properties, such as coming to power, that are not just aggregates of the people in the party because they depend on interactions between the members of the party and interactions between the party and other social groups. Similarly, a key aspect of a solution to the brain–mind problem is appreciation of how psychological properties such as representation, emotion, and even consciousness emerge from neural mechanisms.

Conclusion

In sum, the following principles describe the key relations between the cognitive sciences, which deal primarily with individual thinkers, and the social sciences, which deal primarily with group action: (a) the social does not reduce to the cognitive or the cognitive to the social; (b) explanation of social and cognitive phenomena will require attention to the mechanisms at both levels and to the interactions among them; and (c) social groups can have emergent properties that are neither reducible to nor independent from the psychological properties of their members but rather emerge from them as the result of cognitive and social mechanisms. Philosophy can serve a valuable role by helping cognitive and social scientists make the conceptual and methodological connections that should someday lead to an integrated cognitive social science.

Paul Thagard

See also Artificial Intelligence; Classical Computationalism, Connectionism, and Computational Neuroscience; Emergence; Emergence and Social Collectivism; Group Mind; Holism, in the Social Sciences; Mechanism and Mechanismic Explanation; Mind–Body Relation; Reductionism in the Social Sciences; Social Constructivism; Social Neuroscience

Further Readings

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. New York, NY: Routledge.

Bermudez, J. L. (2010). *Cognitive science: An introduction to the science of the mind*. Cambridge, England: Cambridge University Press.

Boden, M. (2006). *Mind as machine: A history of cognitive science*. Oxford, England: Oxford University Press.

Bunge, M. (2003). *Emergence and convergence: Qualitative novelty and the unity of knowledge*. Toronto, Ontario, Canada: University of Toronto Press.

Elster, J. (2007). *Explaining social behavior.* Cambridge, England: Cambridge University Press.

Glimcher, P. W., Camerer, C., Poldrack, R. A., & Fehr, E. (Eds.). (2009). *Neuroeconomics: Decision making and the brain*. Amsterdam, Netherlands: Elsevier.

Hedström, P., & Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Psychology, 36,* 49–67.

Kaldis, B. (Ed.). (in press). *Mind and society: Cognitive science meets the social sciences* (Synthese Library: Studies in Epistemology, Logic, Methodology, and Philosophy of Science). Dordrecht, Netherlands: Springer.

McCauley, R. N. (2007). Reduction: Models of cross-scientific relations and their implications for the psychology-neuroscience interface. In P. Thagard (Ed.), *Philosophy of psychology and cognitive science* (pp. 105–158). Amsterdam, Netherlands: Elsevier.

Nisbett, R. (2003). *The geography of thought: How Asians and Westerners think differently . . . and why.* New York, NY: Free Press.

Smith, E. E., & Kosslyn, S. M. (2007*). Cognitive psychology: Mind and brain*. Upper Saddle River, NJ: Pearson Prentice Hall.

Thagard, P. (2005). *Mind: Introduction to cognitive science* (2nd ed.). Cambridge: MIT Press.

Thagard, P. (2010). *The brain and the meaning of life*. Princeton, NJ: Princeton University Press.

Turner, M. (2001). *Cognitive dimensions of social science*. Oxford, England: Oxford University Press.

Collective Agents

Collective agents are groups of agents who act together. We speak of groups of agents as doing things and of what they do as collective actions. Collective agents do things both intentionally and unintentionally. Electing a new president is something that citizens do intentionally, but poisoning the environment is something that they do unintentionally. The minimal social unit consists of a group of agents who do something together intentionally. Understanding collective agents and collective agency is therefore central to understanding social reality.

Collective agents sort into two broad categories. Plural group agents are picked out using plural referring terms (e.g., "we," "they," "the President's men," etc.), and their membership is invariable. Singular group agents are picked out using grammatically singular referring terms (e.g., "the Paris Mob," "the Supreme Court," "the British Parliament," "General Motors Corporation," etc.). They can survive changes in their membership and could have had different membership from their present one. These include institutional agents—corporations, governments, armies, clubs, and so on—defined in terms of a structure of institutional roles for individuals, as well as more informal groups such as mobs and crowds.

There are two main questions about collective agents. The first question is whether to understand collective agency we must admit the existence of irreducibly group agents over and above individual agents. Individualists argue that talk of collective agents can be understood entirely in terms of individual agents. Although individualists do not deny that there are groups of agents, they deny that groups are themselves agents. Anti-individualists, in contrast, maintain that collective agency requires (at least in some cases) a group agent over and above the individual agents who constitute it. The second question is how to understand the psychology of group action. Two subquestions arise. First, if we must admit irreducibly group agents, must we attribute to them psychological states, such as desire, belief, and intention; if so, how can we make sense of this? Second, what is the difference between the intentions with which agents act as individuals, I-intentions, and the intentions of agents who participate in intentional group behavior, we-intentions?

Motivations for anti-individualism are divided into three sorts.

The first is the claim that ordinary discourse about collective action entails that there are genuine group agents. The referent of "I" in "I rowed the boat ashore" is the agent of the action expressed. Similarly, it seems, the referent of "We" in "We rowed the boat ashore" is the agent of the action expressed in it. In addition, many actions seem to be things that no individual can be said to have done. When three men lift a piano, no one is the agent of the lifting. A joint declaration of war by Congress cannot be performed even in principle by a single individual. Furthermore, while plural group agents are individuated by their members, singular group agents (e.g., the U.S. Congress) survive changes in their membership, act over periods longer than the life span of any of their members, and could have had different members than they in fact have had.

The second is the claim that we must admit group agents as such to make sense of our practices in assigning praise and blame for collective actions. We seem often to attribute to groups responsibility for what they do (in the form of praise or blame) when it is not clear that any of the individuals who make up the group are candidates for the kind or degree of responsibility in question. For example, in a stoning, no one person who participates is responsible for the death. Or a government may be responsible for what one of its agents does or fails to do, though the agent is excused because of personal obligations that the state does not have.

The third is the claim that some institutions, such as corporations, are complex enough to have the functional organization of an agent. If functional organization is sufficient for being an agent and corporations have the relevant functional organization, it would follow that they were agents in their own right.

Individualism is motivated by the view that the mechanism of joint action lies in the joint expression of individual agency, so that no superagent should be required in order to understand what groups of agents bring about. Individualists can either accept an error theory about ordinary discourse or argue that, on analysis, discourse about group actions does not commit us to group agents but only multiple individual agents contributing to bringing about events. Plausibly, for example, the surface form "We rowed the boat ashore" is misleading. It admits of a

distributive and collective reading. On the distributive reading, we interpret it to mean "Each of us is such that he is an agent of an event, which is a rowing of the boat to shore." We get the collective reading by reversing the order of "an event" and "each of us": "An event is such that each of us is an agent of it, and it is a rowing of the boat to shore." A similar, though more elaborate, strategy would be required for discourse about singular group agents. With respect to the argument from the moral autonomy of group agents, individualists must argue that all moral praise and blame must be distributed over individual members of the groups that act; for example, in a stoning, while no one person is causally responsible for the killing, it might be said that all are equally morally responsible for it, because they act in concert with others with the joint aim of killing the victim. Finally, with respect to the functionalist argument, individualists may charge, even laying aside doubts about the adequacy of a functionalist theory of mind, that the case has not been made that any organization does have the functional organization of a person. The requirement that in attributing any propositional attitude to an agent many supporting attitudes and capacities need to be attributed means that the burden on the functionalist is very heavy. For example, it may seem innocent enough to think of the corporation as such intending to paint its headquarters red, but to possess the concept of redness requires visual experience, for which the corporation, as opposed to the agents who realize it, clearly does not possess the capacity.

Kirk Ludwig

See also Action, Philosophical Theory of; Agency; Collective Intentionality; Collective Moral Responsibility; Group Mind; Holism, in the Social Sciences; Individualism, Methodological; Institutions as Moral Persons; Searle and the Construction of Social Reality; Social Ontology, Recent Theories of

Further Readings

Bratman, M. (1999). *Faces of intention: Selected essays on intention and agency.* Cambridge, England: Cambridge University Press.

Copp, D. (2006). On the agency of certain collective entities: An argument from "normative autonomy." *Midwest Studies in Philosophy: Shared Intentions and Collective Responsibility, 30,* 194–221.

French, P. A. (1979). The corporation as a moral person. *American Philosophical Quarterly, 16,* 207–215.

Gilbert, M. (1992). *On social facts.* Princeton, NJ: Princeton University Press.

Kutz, C. (2000). *Complicity: Ethics and law for a collective age.* Cambridge, England: Cambridge University Press.

List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents.* Oxford, England: Oxford University Press.

Ludwig, K. (2007). Collective intentional behavior from the standpoint of semantics. *Noûs, 41*(3), 355–393.

Miller, S. (2003). Social action: A teleological account. *Australasian Journal of Philosophy, 81,* 300–301.

Schmitt, F. (Ed.). (2003). *Socializing metaphysics.* Lanham, MD: Rowman & Littlefield.

Searle, J. (1995). *The construction of social reality.* New York, NY: Free Press.

Tuomela, R. (2007). *The philosophy of sociality: The shared point of view.* Oxford, England: Oxford University Press.

Collective Emotions

Among social phenomena that raise important philosophical as well as social-scientific issues is the phenomenon of collective emotions. The main purpose of this entry is to present a classification of collective emotional phenomena in terms of their degree or level of *collectivity,* using existing theoretical concepts of collective emotions as examples of different types of collective emotions. The functions of collective emotions in social groups will also be highlighted.

Collective emotional phenomena have been conceptualized in many ways in philosophy and the social sciences. In addition to "collective emotion," notions such as "group emotion," "shared emotion," "collective effervescence," and "emotional climate" have been used in different discourses. These notions are partly overlapping, but there is no wide interdisciplinary agreement on the definition of "collective emotions." Therefore, this concept will here be used as an umbrella term for collective phenomena that are characterized by *intentional aboutness* and *affective experiential quality,* as distinct from other affective phenomena such as moods or feelings, which lack the intentionality of emotions.

Even if intentionality and affectivity are phenomenologically inseparable in the emotional content,

we can distinguish the question about the intentional structure of collective emotions from the question about their affective quality. Intuitively, collective emotions are collective in both respects—their intentional structure is collective in some robust sense, and their affective experience is collective in the sense of being shared by several individuals. However, these aspects of collective emotions may require different kinds of explanation. The subject, object, and content of emotion are important for the collectively intentional structure of emotions, whereas physical proximity of individuals and psychological and biological mechanisms of various kinds, such as attentional deployment, facial mimicry, neural mirroring, behavioral entrainment, and emotional contagion, are relevant for the emergence of collective affective experiences.

The Intentional Structure of Collective Emotions

Following Margaret Gilbert, we can divide theoretical accounts on the *intentional structure* of collective emotions into three types: weakly aggregative accounts, strongly aggregative accounts, and collective subject accounts. The subjects of the first two types of collective emotions are individuals, while the main difference between these emotions lies in their content. Weakly aggregative collective emotions are about the subject's own actions or events that matter to a subject on the basis of his or her private concerns, whereas the content of strongly aggregative collective emotions is associated with the subject's *group membership* and the shared concerns that come with it. Finally, collective subject accounts suggest that even if there are, ontologically speaking, no group minds but only individual minds, collective emotions nevertheless have a "collective subject" in some robust sense.

Weakly aggregative collective emotions are collective in a weak sense of being based on overlapping private concerns of individuals. These concerns may be about survival, security, attachment, health, wealth, happiness, and so on. An example of a weakly aggregative collective emotion is the panic of shareholders about the fall of stock prices. Individual shareholders share the same type of emotion about the same event. However, the content of their similar emotions is individual, because each shareholder is concerned about his or her own wealth. A theoretical concept of this type is J. M. George's "group affective tone." Here, it is sufficient that an aggregate of individuals experience the same type of emotion, such as fear or anger or joy, in a group context, such as at a bus stop or workplace, with mutual awareness that others are feeling the same.

Strongly aggregative collective emotions are experienced by a subject in the role or position of a group member, with mutual awareness that other group members are feeling the same type of emotion. The content of emotion is collective, because the emotion is about an event that matters to the individual on the basis of his or her membership in a particular social group or category, such as a political party, sports team, religious sect, artistic ensemble, workgroup, or ethnic group. Empirical and philosophical theories differ from each other in terms of how they understand the required group membership. Empirical theories, such as the *Intergroup Emotion Theory* of D. M. Mackie and E. R. Smith, emphasize the psychological mechanism of group identification, whose intensity determines the relative importance of different groups for individuals. However, from a philosophical perspective, the intensity of identification does not distinguish between those groups whose membership is founded on individuals' private commitment to the group or its ethos—certain constitutive goals, values, norms, or beliefs of the group on the one hand—and those groups whose membership is founded on the group members' joint commitment to the group or its ethos on the other (as in Raimo Tuomela's theory). From a normative point of view, the latter kind of collective emotions are more robustly collective than the former kind of collective emotions. The "membership emotions" Margaret Gilbert identifies provide an example of the more robust type of strongly aggregative collective emotions.

Accounts in terms of "collective or plural subjects" argue that the subject of collective emotions is in some sense a "we" rather than a mere aggregate of individuals who ontologically constitute the specific "we." Margaret Gilbert suggests that members of a plural subject can jointly commit themselves to feeling particular emotions as a body. The object of commitment is an evaluative judgment that for Gilbert constitutes the content of emotion, whereas feelings are only contingent concomitants of emotions. Therefore, having a collective emotion does not require experiencing it but merely acting and talking in ways that are consistent with the evaluative content of the emotion. If, for instance, a group of

people have jointly committed themselves to feeling guilt, then they have jointly committed themselves to judging the object of their emotion, a certain action of theirs, as morally wrong, as well as to talking and behaving in accordance with this judgment.

H. B. Schmid rejects Gilbert's plural subject account, because it undermines the affective dimension of emotions. However, he retains the idea of a collective subject, which in his model emerges as a phenomenological frame in which individuals interpret their feelings as "our" feeling. A phenomenological fusion of feelings requires that individuals share the same concern, such as survival, attachment, or welfare, but their emotions need not have the same object. Schmid also suggests that the feelings of individuals may differ from each other in terms of their intensity and quality without undermining the possibility of a phenomenological fusion.

The categories of weakly and strongly aggregative collective emotion relate to real emotions at various levels of collectivity. A phenomenological fusion of feelings also seems possible, but it need not alone indicate a high level of collectivity because such fusion may emerge even in the context of weakly aggregative collective emotions. Instead, the plural subject account of Gilbert does not appear to refer to any genuine collective emotion, for it is not psychologically possible to summon an emotion by means of a joint commitment to feeling it. Nevertheless, Gilbert's plural subject account of collective emotions is plausible in the sense that it refers to feeling rules—rules of appropriateness for distinct emotions—which individuals can, in some circumstances, produce by jointly committing themselves to such rules. Feeling rules are important in the regulation of emotional states and expressions, but they are emotions themselves. Strongly aggregative collective emotions that are felt for reasons emerging from a jointly committed group membership represent then the highest level of collectivity as far as the intentional structure of collective emotions is concerned.

Collective Affective Experience

The emergence of collective affective experiences requires the copresence of individuals, either physical or virtual. The key to such experiences is an extensive *synchronization* of several individuals' emotional responses: their physiological changes, facial expressions, action tendencies, and subjective feelings. Several psychological and biological mechanisms contribute to such synchronization. These mechanisms include attentional deployment, facial mimicry, neural mirroring, behavioral entrainment, and emotional contagion. The sociologist Randall Collins maintains that the synchronization of individual emotional responses typically occurs in the context of interaction rituals, such as songs, dances, chants, games, and other activities, in which the participants are separated from others; they share a common emotional mood, their attention is focused on the same object or activity, and they are mutually aware of sharing the same mood and focus of attention with each other. The result of extensive synchronization of individuals' emotional responses is collective effervescence, "a high degree of absorption in emotional entrainment."

The Functions of Collective Emotions in Social Groups

Collective emotions have several functions in the emergence and maintenance of social groups. First, collective emotions contribute to the emergence of social groups as well as to the formation of their goals, values, and intentions—a certain group ethos. For instance, shared anger at oppression or shared guilt about collective wrongdoing to a third party contributes to the emergence of social and political movements. Second, collective emotions motivate expressive and purposive behavior in support of the group ethos. Shared anger motivates protests, and shared guilt motivates apologetic and reparative actions toward a violated outgroup, for instance. Third, collective emotions manifest and reinforce group members' commitment to the group and its constitutive ethos. Group members feel shared emotions when their group ethos is affected favorably or adversely, thus *experiencing* their being "in the same boat" with each other. Fourth, collective emotions affectively "glue" group members to each other as fellow group members. Collective emotional experiences are intrinsically pleasant and rewarding as they involve feelings of solidarity, which foster affective bonds, cooperative ties, and group loyalty, especially if the shared emotions are hedonically positive rather than negative. There is some evidence that strongly collective emotions are capable of serving these functions more efficiently than weakly collective emotions, but this hypothesis requires further support

from empirical research that would pay closer attention to the various types of collective emotions.

Mikko Salmela

See also Affective Intelligence in the Social Sciences; Collective Agents; Collective Goals; Collective Intentionality; Commitment; Emotions; Group Identity; Group Mind; Joint Attention and Social Cognition; Plural Subjects; Solidarity; We-Mode, Tuomela's Theory of

Further Readings

Collins, R. (2004). *Interaction ritual chains*. Princeton, NJ: Princeton University Press.

George, J. M. (1996). Group affective tone. In M. A. West (Ed.), *Handbook of work group psychology* (pp. 77–93). Chichester, England: Wiley.

Gilbert, M. (2002). Collective guilt and collective guilt feelings. *Journal of Ethics, 6,* 115–143.

Salmela, M. (2012). Shared emotions. *Philosophical Explorations, 15,* 33–46.

Schmid, H. B. (2009). *Plural action: Essays in philosophy and social science* (Contributions to phenomenology: Vol. 58). Berlin, Germany: Springer.

Smith, E. R., Seger, C. R., & Mackie, D. M. (2007). Can emotions be truly group level? Evidence regarding four conceptual criteria. *Journal of Personality and Social Psychology, 93,* 431–446.

Tuomela, R. (2007). *The philosophy of sociality.* New York, NY: Oxford University Press.

Collective Goals

The concept of a "collective goal" is central to several problems relating to collective action. For example, it has been argued that collective actions are brought about when a group of individuals orient themselves around a collective goal. Similarly, a collective intention is characterized—at least in part—by a desire on the part of a group of individuals to bring about a collective goal. To take another example, Margaret Gilbert has influentially argued that the collective goals of a group play an important role in determining the beliefs that are held by the group.

Despite its centrality, there have been relatively few full-length treatments of this concept in the collective action literature. Notable exceptions are Raimo Tuomela and Kaarlo Miller, who emphasize the differences among various kinds of goals, including collective goals. According to Tuomela, a collective goal must satisfy two conditions. First, it must be accepted in the "we-mode"; this means that each person aims to achieve the goal as a member of the group. Second, it must satisfy the "collectivity condition": if the goal is met for anyone in the group, it is met for everyone in the group. These conditions are echoed by other authors, including Christopher Kutz, who argues that for a goal to be collective, there must be a nonempty intersection of states of affairs that satisfy those collective ends.

The clearest way to understand these conditions is by contrasting collective goals with goals that are not collective and with group behaviors that merely resemble those brought about by collective goals. For example, a collective goal is not merely a goal that happens to be shared by more than one agent. So if you and I each have the goal of going to a particular restaurant at noon, we do not thereby have a collective goal to do so. Even though there is a state of affairs in which our goals are satisfied—namely, the state of affairs in which we are both at the restaurant at noon—our goals can be satisfied without being in the "we-mode," that is, without our aiming to meet our goals cooperatively as a group. As Tuomela observes, this distinction between a goal that happens to be shared by all the members of a group and a truly collective goal corresponds to Rousseau's distinction between the "will of all" and the "general will."

To take another example, we may have the goal that a certain table be moved from one end of a room to another. If we are indifferent as to how the table is moved—that is, if we don't care who moves it—then it is not a collective goal. In other words, if the goal can be met by my moving the table by myself, then the goal cannot be collective. In contrast, a truly collective goal that we move the table *together* can be satisfied only if the table is moved as a result of our cooperative action. Such a collective goal would be thwarted if one of us were to move the table alone.

It is also useful to note the differences between meeting a collective goal and merely coordinating our behavior. This contrast is informative because it shows that important conditions that are frequently discussed in theories of collective action are insufficient for characterizing collective goals. For example,

consider the coordinated behavior of driving on the right side of the road in some countries. As individuals, we each have the goal of driving on the right, and our goals are mutually interdependent in the sense that they are justified by our knowledge that others have the same goal. In this respect, the example is different from the case where we each want the table to be moved; in the latter case, the rationality of my goal did not depend on your having the same goal. But in this case, it would be irrational for me to have the goal of driving on the right side of the road if nobody else had the same goal. But despite this interdependence, it is generally accepted that there is no collective goal to drive on the right side of the road.

Furthermore, collective goals share a particular self-referential property with intentions. For example, my intention to break a window is clearly not satisfied if you throw me through the window, thereby causing me to break it. Similarly (to adapt one of Michael Bratman's examples), our collective goal of going to New York together is not satisfied if we are both abducted and brought to New York together in the back of a van. In this way, a collective goal is self-referential in the sense that it is met only if the appropriate state of affairs is brought about by virtue of our having the collective goal.

Sara R. Chant

See also Collective Agents; Collective Intentionality; Goal-Directedness; Group Mind; We-Mode, Tuomela's Theory of

Further Readings

Bratman, M. (1993). Shared intention. *Ethics, 104,* 97–113.

Gilbert, M. (1992). *On social facts.* Princeton, NJ: Princeton University Press.

Kutz, C. (2000). Acting together. *Philosophy and Phenomenological Research, 61,* 1–31.

Miller, K., & Tuomela, R. (2001). What are collective goals? In M. Kiikeri & P. Ylikoski (Eds.), *Explanatory connections.* Retrieved from http://www.helsinki.fi/teoreettinenfilosofia/henkilosto/Sintonen/Explanatory%20Essays/miller.pdf

Tuomela, R. (2008). Collective intentionality and group reasons. In H. B. Schmid, K. Schulte-Ostermann, & N. Psarros (Eds.), *Concepts of sharedness: Essays on collective intentionality* (pp. 3–19). Berlin, Germany: Springer-Verlag.

Collective Identity and Cultural Trauma

An important feature of social life is traumatic experiences felt by collective actors when public meanings get shattered or drastically rearticulated. This entry charts the various aspects of cultural trauma and how it relates to groups and in particular to collective identity, discusses the role of emotions, and explains cultural traumas as essentially processes. As well as being destructive of established social or national (and other) identities, traumas may also be conducive to, or productive of (renewed), collective identities.

Cultural trauma is a discursive response to a tear in the social fabric, when the foundations of an established collective identity are shaken by a traumatic occurrence and are in need of re-narration. The process of cultural trauma can be understood as a meaning struggle, where individual and collective actors attempt to define the situation and impose a particular meaning on it. Within this meaning struggle, various individuals and groups construct narratives in which trauma is an organizing notion.

Cultural trauma should be distinguished from the classical and commonsense notions of trauma, which have in common the naturalistic assumption that trauma results from a wound inflicted on the body or the mind through an overwhelming event that imposes itself on a victim. This naturalistic notion roots trauma in direct individual experience, one that leaves an indelible mark that will later cause the appearance of otherwise inexplicable behavior. By contrast, cultural trauma involves discursive practices, collectivities, collective memory, and collective identity.

Definitions

In an insightful discussion of the difference between individually rooted psychological and cultural trauma, Neil Smelser finds one essential difference in the fact that cultural traumas are made, not born. He goes on to define a cultural trauma as "an invasive and overwhelming event" that is believed as undermining a whole culture as such or some of its essential ingredients. Though he acknowledges the discursive aspect of cultural trauma, Smelser grounds the process in an event. Others (e.g., Ron Eyerman) argue that an event is also a discursive

product and distinguish between a traumatic occurrence and a cultural trauma. This is also suggested in Smelser's theory with the insertion of the phrase "believed to undermine," a notion that undercuts any idea that an event could be traumatic in itself.

This proposition opens two vital questions: (1) Can any occurrence or event be made traumatic, so that it is "believed to undermine" an established collective identity? (2) If not, what is it that permits some occurrences to become traumatic in this sense and not other seemingly equally powerful or shocking occurrences? The first question points to the power to create belief. A cynic might argue that given the ultimate power to persuade, one could turn any occurrence into a "trauma." This would be to push the idea that "traumas are made, not born" to its limits. At the other extreme, a strong naturalist or lay trauma account would make the claim that certain events are traumatic in themselves, that is, they are the direct cause of traumatic affect. A more moderate position might argue that certain occurrences create conditions conducive to setting in motion a process of cultural trauma, without necessarily being traumatic in themselves. Such a process will not happen without the aid of mediating, meaning-making forces, such as the mass media and other more specifically defined carrier groups, such as intellectuals, who articulate the suffering and influence the formation and direction of a process of cultural trauma. However, not all or any interpretative frame will "fit" or make sense; there must be some relation, real or perceived, to some referent—an occurrence, experience, or event—which itself appears "always there."

Cultural Traumas as Processes

Cultural traumas are not things but *processes* of meaning making and attribution, a contentious contest in which various individuals and groups struggle to define a situation and to manage and control it. It can also be added that these forces are unlikely to create a trauma out of nothing; there is likely to be some powerful, shocking occurrence that creates the possibility, providing the opportunity to mobilize opinions and emotions. There are thus two sides to a cultural trauma: an emotional experience and an interpretative reaction. Shocks arouse emotion by breaking everyday routines (behaviors as well as cognitive frameworks) and as such demand interpretation, opening a discursive field where well-placed individuals can play a determinative role. In modern societies, access to the mass media is significant in this process. The polarity between perpetrator and victim is what distinguishes cultural trauma as a discursive process. In this sense, cultural trauma is a contentious struggle to define the meaning of a shocking, unforgettable occurrence framed by a dichotomy between perpetrator and victim, where the foundations of a *collective identity* are shaken.

Traumas and Public Emotions

Yet cultural traumas are more than a struggle between competing actors to define a situation, to distinguish between perpetrator and victim, and to identify the nature of the pain. That would limit the process to instrumental or strategic interaction. Cultural traumas are a response to deeply felt emotions that are publicly expressed and represented in this very same process, which implies an expressive and communicative aspect that makes claims to authenticity and sincerity connected to *collective identity* and *memory.* Anything connected to identity falls within the realm of the sacred and as such is bound up with powerful emotions. There are deeply rooted emotions and scripted identities to be drawn upon in such a situation.

One can think, for example, of ethnic or national identities that may lie under the surface but that can be mobilized in the face of a shocking occurrence, such as a political assassination. This was the case with the assassination in 1914 of Archduke Ferdinand of Austria, and the process started by this occurrence spiraled into a world war that fundamentally altered the political geography of Europe.

"Believe to Undermine or Overwhelm" Collective Identities

We can now return to the question of how it is that a collective comes to *believe* that an occurrence "overwhelms" their already established collective identity. Collective identities are rooted in beliefs that are maintained in everyday life through routine practices. Routines provide confirmation and security in that they allow beliefs to be taken for granted, to be in a sense forgotten. An example of Eyerman's may help illustrate this. One of the first public statements made by the Swedish prime minister after the assassination of Foreign Minister Anna Lindh in 2003

was "This is an attack on our democratic society." That Sweden is a democratic society is a fundamental belief and value grounding modern Swedish collective identity. This belief is not only taught in schools but is also bound up with routine practices such as voting every several years to elect the government. The fact that Sweden is a "democracy" is normally taken for granted. The murder of Anna Lindh was shocking not only because she was a well-known member of the collective but also because she was a political figure, a representative of that democratic process. At least for the prime minister, another representative figure, her murder presented a threat to that fundamental value/identity, as well as to the political process that underpins it. It was thus important for the security and the stability of that identity not only that the murderer be caught, which was a police matter, but also that the rules and procedures of the political system that would guarantee stability were immediately put on display. This was even clearer in the assassination of Prime Minister Olof Palme in 1986, where the same radio bulletin that announced the murder to the public also confirmed that the government was already meeting to ensure the succession of his replacement. Since shocking events like political assassinations, or the explosion in the center of Oslo, Norway, and the subsequent massacre at nearby Utøya Island on July 22, 2011, break everyday routines and can call into question fundamental taken-for-granted beliefs that ground individual and collective identity, it is important that those in positions of authority act quickly to reaffirm those basic identities. To act in this manner is one way of assuring, or attempting to ensure, that the shock caused by a political assassination will be contained and limited to an institution or set of institutions, in this case that of politics and law enforcement, and not involve the society at large. Failure to do so, or failure of such performances of authority, increases the risk that a shocking occurrence will become a cultural trauma.

This returns one to Smelser's discussion of cultural trauma—of shocking occurrences as being "potentially" traumatic—which must be successfully "endowed with negative affect" in order to be fully realized. Central in the meaning struggle and thus to the making of a cultural trauma is the successful attribution of a term such as *national tragedy, national shame,* and *national catastrophe* and its acceptance by a significant part of the collective. This is what is meant by the phrase *come to believe to undermine*: that a significant number, who can say how many, of members of a collective come to believe that the shock is a "national tragedy" and has undermined the fundamental values that have defined the collective. This is what turns or transforms a shock into a cultural trauma and leads to Smelser's (2004) more formal definition:

> A cultural trauma is a memory accepted or given credence by a relevant membership group and evoking an event or situation which is (a) laden with negative affect, (b) represented as indelible, and (c) regarded as threatening a society's existence or violating one or more of its fundamental cultural pre-suppositions. (p. 44)

The Collective

Even when speaking of a collective, one must ask "trauma for whom?" Imaginary collectivities, such as nations or ethnic groups, are rarely unified or univocal. One effect of a traumatic occurrence can be to provide a sense of coherence and collectivity, even if this is also imaginary and temporary. The attack on New York's World Trade Center in 2001 appeared to unify the American nation into an emotional collective, producing ritual practices that helped sustain it, just as the phrase "9/11" is meant to evoke and signify shared experience and collective understanding. But digging beneath that ephemeral surface, in large part facilitated through mass-mediated representations, one would undoubtedly find individuals and even groups that would dissent from that feeling. In this sense, a traumatic occurrence has the potential to both unify and divide, to create both insiders and outsiders.

This potential must, however, be realized, and it is here that what Jeffrey Alexander, following Max Weber, called *carrier groups* play an important role. Such groups articulate and represent trauma, making it available for communication and shared understanding. They help transform emotional response into words and images that can be dispersed and remembered. Artists, writers, journalists, and political and religious leaders are important social categories in this articulation and play an important role in the trauma process. Potential carrier groups are broader than these professional categories, however. Carrier groups, such as family members and friends, for example, can be preexisting or may form

in response to a particular traumatic occurrence, even while it is likely that professional categories, like those mentioned above, will become significant agents within the process of cultural trauma. Carrier groups not only are central to the making of cultural trauma, they are important in its continued affect. Carrier groups are bearers of memory and can, through their actions, elicit emotional recall in distant others.

Cultural trauma is thus a process through which collectivities are articulated, formed, and re-formed in light of traumatic occurrences. Traumatic occurrences and the cultural trauma process need not only have negative outcomes for all concerned; they can create as well as shatter collectivities.

Ron Eyerman

See also Collective Emotions; Collective Memory; Collective Values; Group Identity; Identity, Social; Personal Identity and Trauma

Further Readings

Alexander, J. (2004). Toward a theory of cultural trauma. In J. Alexander, R. Eyerman, B. Giesen, N. J. Smelser, & P. Sztompka (Eds.), *Cultural trauma and collective identity* (pp. 1–30). Berkeley: University of California Press.

Alexander, J., Eyerman, R., Giesen, B., Smelser, N., & Sztompka, P. (Eds.). (2004). *Cultural trauma and collective identity.* Berkeley: University of California Press.

Bartmańński, D., & Eyerman, R. (2011). The worst was the silence: The unfinished drama of the Katyn massacre. In R. Eyerman, J. Alexander, & E. Breese (Eds.), *Narrating trauma* (pp. 237–266). Boulder, CO: Paradigm.

Eyerman, R. (2001). *Cultural trauma slavery and the formation of African American identity.* Cambridge, England: Cambridge University Press.

Eyerman, R. (2008). *The assassination of Theo van Gogh*. Durham, NC: Duke University Press.

Eyerman, R. (2011). *The cultural sociology of political assassination*. New York, NY: Palgrave/Macmillan.

Eyerman, R., Alexander, J., & Breese, E. (2011). *Narrating trauma*. Boulder, CO: Paradigm.

Giesen, B. (2004). The trauma of the perpetrators: The holocaust as traumatic reference of German national dentity. In J. Alexander, R. Eyerman, B. Giesen, N. J. Smelser, & P. Sztompka (Eds.), *Cultural trauma and collective identity* (pp. 113–154). Berkeley: University of California Press.

Smelser, N. (2004). Psychological and cultural trauma. In J. Alexander, R. Eyerman, B. Giesen, N. J. Smelser, & P. Sztompka (Eds.), *Cultural trauma and collective identity* (pp. 31–59). Berkeley: University of California Press.

Collective Intentionality

Explanations of human agency, both in everyday contexts and in the context of social scientific research, often appeal to intentions, beliefs, and other mental states of individuals. But humans act together, forming social groups that engage in a variety of activities. *Collective intentionality* is a subfield of social philosophy that attempts to answer questions regarding the nature of social groups, group action (collective agency), and whether a group can be held responsible for its actions (collective responsibility). This entry focuses on the philosophical debates surrounding collective intention and collective belief as they play a dominant role in discussions of group agency and responsibility.

Philosophical Debates About Collective Intention

Consider the attribution of intention to groups such as teams, corporations, committees, and organizations. One might attempt to explain the collective intention expressed by the phrase "We intend to win the soccer game!" by reducing it to the sum of individual intentions (to win) had by members of the group. On this view, a group intends to X if and only if each member (or perhaps a majority of the members) intends to X, or each individual intends to do his or her part in bringing about X and to do so under conditions of common knowledge. But the existence of individual intentions to do one's part even under conditions of common knowledge has seemed, to some philosophers, to be insufficient to explain collective action and our attributions of collective intention. According to John Searle, for instance, collective intentions cannot be explained in terms of individual intentions because the collective intentions are primary. Even if each member intends to do his part to bring about X, these individual intentions seem to derive from the collective intention. It is because there is a collective intention that individuals are able to form individual intentions to

do their part. But are collective intentions the intentions of some supra-individual or group mind?

Like many others working in the area of collective intentionality, Searle is committed to a form of ontological individualism. Ontological individualism is the view that individuals are the primary and sole units of agency. There are no collective minds or supra-agents that are intentional agents in any real sense. Because there is no supra-agent that is the subject of the collective intention and all intentionality is to be found in the heads of individual human beings (and animals), Searle posits the existence of "we-intentions." We-intentions are states of individuals, not groups. Just as an individual can intend from the first-person perspective, one can also intend from the third-person-plural perspective. From we-intentions, one derives specific intentions of the form "I intend to do my part to bring about X." Searle offers a similar analysis of collective belief. Collective beliefs are not the beliefs of a group but the beliefs of group members, individuals have the ability to believe in the "we-mode."

For some, this does not sufficiently explain joint action and our attributions of intention to groups. There is something about collective intentions that coordinates individual, independent actions into a joint action. But isolated we-intentions do not, by themselves, seem to be enough to direct and coordinate the individual intentional actions that make up joint action. This suggests that the individual intentional states that underlie collective intentions should be suitably interrelated.

Michael Bratman provides an account of collective intention in terms of the intentions of the individual participants and their interrelations. Bratman's account also preserves a form of ontological individualism. However, on his view, while collective intentions are not the intentions of some supra-agent, they are not located in the heads of individuals either. Rather, shared intentions are states of affairs that consist in a set of individual intentional states and their interrelations. On Bratman's account, when individuals share an intention, each individual has the intention of the form "I intend that we X," and these intentions are formed because of and by way of the intentions of others. Bratman's analysis also requires that subplans of individuals (the means by which they plan to do their part to bring about X) do not conflict or fail to mesh. In recent work, Bratman has developed this account to explain more complicated forms of collective action and introduced the notion of shared value.

Bratman and Searle's analysis of collective intentions has been subject to the following objection: Normally, one cannot intend to do what one does not have control over. Consider again the following collective intention ascribed to a soccer team: "*Purple Storm* intends to win the game." Since no one individual has control over the actions of his teammates, it is difficult to explain how any individual could have the intention "We intend to win." An individual might have the intention to do her part by scoring a goal or assisting others. The actions required to fulfill these intentions are within her control. But she cannot intend that others score goals and make assists. She can wish and hope that they do so. But she cannot intend that they do so. The individual who does the intending is normally the same individual who fulfills the intention. If an individual can only intend what is within her control, then contrary to Searle and Bratman, individuals cannot have we-intentions or intentions of the form "I intend that we X."

The distinction between intending *to* and intending *that* has been appealed to in order to respond to this criticism. Although an individual cannot *intend to* get married (because such an act requires the action of another that is out of one's control), she can *intend that* she and her partner get married. In a similar vein, Christopher Kutz argues that joint action can be explained by appeal to a special type of intention called a "participatory intention." These are not intentions *to do* something but individual intentions *to do one's part* in a collective action. Seumas Miller avoids these difficulties by explaining collective action in terms of collective goals or ends, rather than appeal to the notion of intention.

Philosophical Debates About Collective Belief

We often attribute beliefs to groups such as corporations, and social scientific research might appeal to a group's beliefs in order to explain its behavior. In some cases, such ascriptions will identify beliefs held in common. That is, to say that when group G believes that *p*, it might mean that all or most of the members of G believe that *p*. But in other cases, the ascription of group belief might attempt to identify something beyond a shared belief. Margaret Gilbert, for instance, argues that there are cases when we

ascribe a belief to a group even when no individual member of the group has the belief being ascribed to the group, or at least such cases are conceivable. Thus, for a group to be said to believe that *p,* it need not be the case that all or most of the members believe that *p*. In some cases, group members might accept that *p* is the belief of the group and act in accordance with that belief as group members. Or as Gilbert describes it, there may be a joint commitment to believe that *p* as a body or as a group. When individuals jointly commit to believing that *p* as a group, there is a plural subject that believes that *p*. Joint commitments are the glue that forms plural subjects, and joint commitments cannot, according to Gilbert, be reduced to individual commitments to believe that *p*.

Raimo Tuomela offers an extensive taxonomy and analysis of different types of collective belief. In some cases, our ascriptions capture shared we-beliefs. In the case of we-beliefs, each member believes that *p,* and it is common knowledge that each member believes that *p*. But in some cases, our ascriptions of collective belief move beyond we-beliefs. Tuomela appeals to the distinction between the I-mode and the we-mode in order to capture a different form of collective belief. Just as we can personally believe something (the I-mode), we can believe a proposition from the we-mode. Tuomela has also developed an influential account of corporate beliefs. Collective beliefs in the case of large groups such as corporations can be explained by relying on the notions of rule-based social positions, tasks, and positional beliefs. Positional beliefs are the views held by a person qua position holder, as one with decision-making authority. They are views that one accepts as a basis for action, without necessarily being views that one personally believes. Positional views, then, need not be truth related. We may accept false beliefs and adopt positional views that we know to be false.

The accounts considered thus far attempt to explain collective intention and collective belief in terms of other concepts such as acceptance, joint commitment, and we-beliefs. According to most accounts, collective intentionality is not the intentionality had by groups but a special form of intentionality had by individuals as group members. But one might argue that collective intentions and beliefs are to be understood as states of the group *itself*. There are two ways to go about this. One might argue that intentional states such as belief and intention should be understood functionally: Though collective intentions and beliefs are realized by states of individuals and their interrelations, they are had by the group because they play the same functional role within the group that individual intentions and beliefs play within the individual mind. Another approach would be to adopt a form of instrumentalism: Intentions and beliefs are not states of the brain or mind at all but are mere explanatory constructs that are useful in explaining and predicting the behavior of certain kinds of systems. Deborah Tollefsen has argued that instrumentalism lends itself nicely to the idea that groups, too, can be the subject of intentional state ascription, such as belief and intention. To the extent that one takes the intentional stance toward certain types of groups and doing so helps explain and predict collective action, groups themselves "have" beliefs and intentions.

The Role of Collective Intention and Belief

Theories of collective intention and belief have had an impact on a number of philosophical and social scientific issues and debates. Searle's theory of collective belief and intention serves as the foundation for a theory of social facts. Collective intentions confer status functions on artifacts and turn them into social facts. The bits of paper we call money serve a particular role in our lives. Money is money because we accept it and intend it to serve a particular role. Others have explored the role that collective intentionality, either collective intentions or beliefs, plays in political philosophy and moral theory. Gilbert, for instance, argues that her account of collective intentionality provides a richer account of political obligation and social rules than competing theories. Tuomela has developed accounts of rule-governed social practices and fads by appeal to collective intentions and beliefs. Fields as diverse as cognitive ethology, evolutionary psychology, developmental psychology, jurisprudence, economics, and sociology have appealed to work in collective intentionality to explain both human and animal behavior.

Deborah Perron Tollefsen

See also Collective Agents; Collective Goals; Collective Moral Responsibility; Collective Values; Group Beliefs; Group Mind; Institutions as Moral Persons; Plural Subjects; Searle and the Construction of Social Reality; Social Facts; Social Institutions; Social Ontology, Recent Theories of; We-Mode, Tuomela's Theory of

Further Readings

Bratman, M. (1993). Shared intention. *Ethics, 104,* 97–113.
Bratman, M. (1999). *Faces of intention.* Cambridge, MA: Cambridge University Press.
Gilbert, M. (1989). *On social facts.* New York, NY: Routledge.
Gilbert, M. (2003). The structure of the social atom: Joint commitment as the foundation of human social behavior. In F. Schmitt (Ed.), *Socializing metaphysics: The nature of social reality* (pp. 39–64). Lanham, MD: Rowman & Littlefield.
Miller, S. (2001). *Social action.* Cambridge, MA: Cambridge University Press.
Searle, J. (1990). Collective intentions and actions. In P. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 401–416). Cambridge, MA: Bradford Books/MIT Press.
Searle, J. (1995). *The construction of social reality.* New York, NY: Free Press.
Tollefsen, D. (2002). Organizations as true believers. *Journal of Social Philosophy, 33,* 395–411.
Tuomela, R. (1992). Group beliefs. *Synthese, 91,* 285–318.
Tuomela, R. (2005). We-intention revisited. *Philosophical Studies, 125,* 327–369.

Collective Memory

The entry introduces the meaning of the term *collective memory,* traces its origins and explains its anti-reductionist or anti-individualist nature in social ontology, and goes on to review recent formulations of it.

Meaning

Collective memory denotes the totality of a collective entity's mnemonic practices. While personal memories relate to an individual's experiences with the past, these experiences—to the extent that they are mnemonically re-created in the present at all—represent a small slice of past social reality, the rest of which is forgotten. Collective memory transcends the sphere of the individual in that individual experiences with the past relate to the experiences of others so as to render the past meaningful to both the individual and the collective; acts of remembrance (and forgetting) are collective performances, and the impact of remembrance is collective. The late 20th and early 21st centuries are said to have experienced an enduring "memory boom" centered on the Holocaust, marked not only by an increasing number of memorials but also by a plurality in memory.

Origin

When the French sociologist Maurice Halbwachs coined the term *collective memory* to designate a social phenomenon that emerges from social interaction, he stood in the tradition of Durkheimian sociology. Halbwachs's mentor Émile Durkheim noted that collective representations of the world, including those of the past, have their origins in interactions in collective entities from the beginning and *cannot be reduced* to the contributions of individuals. Remembered events and experiences are hardly ever constituted by individuals divorced from other individuals or from a social framework that constitutes a generative milieu for the events and experiences. Rather, they result from *collective acts,* and the memory of events and experiences is reconstituted and affirmed in collective acts of remembrance. Entities that relate to the past in this way can vary in scope from small groups to entire nations and even the global community, as can the entities affected by remembrance; that is, the impact of collective memory is social as well as individual.

Developments

In conceptualizing collective memory, scholars such as Jörg Assmann differentiate between communicative memory and cultural memory. *Communicative memory* is tied to living or recent generations. Traditions are passed on mostly orally and thus depend on communicative transmission among persons. Communicative memory tends to be more fluid, less formal, and more tied to daily life than cultural memory. *Cultural memory* provides a long-term storage function for the types of knowledge that help shape a society's collective identity and from which individuals derive their cultural capital. The distinction between communicative and cultural memory points to the importance of forms of literacy, communication media, collective time horizons, and the social functions of collective memory: Oral cultures draw on and perpetuate living memory; elite literacy is tied to the creation and distribution of manuscripts and serves to recover lost wisdoms; mass literacy and the availability of print media allow the reconstruction of a past distinct from the present; and literacy tied to electronic media in the

digital age facilitates the deconstruction of representations of the past.

Literacy constitutes one element in the cultural settings of action that frame how, where, and when past events are remembered. At a macro-social level, such settings may be part of larger memory regimes. Memory regimes reflect the ways in which national histories, politics, and culture have shaped commemorative practices, and over the course of the past century they have varied greatly, but the engagement with the past is also often intensely local. That is, whether the past is remembered at all, and to what extent, is an issue that relates not only to the germinating effects of culture but also to the activities of local memory entrepreneurs and agents as well, seeking to imbue selective elements of the past with cultural significance.

Some scholars note that the interplay of national and regional memory regimes and local memory agents has taken on a new quality in the digital age, in which collective memory is increasingly constituted via electronic media. Representations of the past, the historian Pierre Nora has argued, are no longer tied to what he calls *milieux de mémoire,* or lived memory that is a part of everyday life and reconstituted in traditional rituals, but are rather embodied in "sites of memory" (*lieux de mémoire*), or any place, practice, object, or concept that serves as a symbol by which a social entity relates to its past. For Nora, sites of memory are not necessarily material, but they are always socially constructed, and he deems them central to national identity formation. Postmodernist thought has taken this argument further, referring to a rapidly changing order, a disbelief in the linearity of time, the prominence of simulations, and constancy and continuity found only in the experience of transience itself, conditions that lend themselves to memories being dominated by images and their reproductions.

A "memory boom" since the late 20th century has led to the emergence of numerous new sites of memory. A considerable number of sites relate to the Holocaust, some entirely virtual. In the decades following World War II, memories of the Holocaust served different functions: In Israel after the Eichmann trial in 1961, they were employed to advocate a need for a strong state that could militarily defend itself in the face of adversity; in the United States, they gradually became part of a more general narrative of success in the American quest to liberate the world from evil; in Eastern Europe, they were suppressed by communist governments or considered part of a larger group of memories of victimhood, lest they detracted from a state-sponsored focus on resistance fighters as antifascist heroes; in Austria, they were quickly set aside in a national attempt to claim the status of first victim of Nazi Germany or became part of a narrative of attributing the Holocaust to Germany in order to absolve the nation from culpability in participating in it; in West Germany, the assumption of responsibility for the Holocaust helped the government reintegrate German society back into the fold of Western nations. More recently, as part of a trend toward the memorialization of trauma, the Holocaust as a historically unique atrocity has become a generalized symbol of human suffering and moral evil. Some see the generation of knowledge about it and the memorialization of its victims as opportunities for humanistic learning in a global memory culture proffered by the formation of cosmopolitan memory.

Lutz Kaelber

See also Collective Agents; Collective Identity and Cultural Trauma; Durkheim's Philosophy of Social Science; Group Mind; Holism, in the Social Sciences; Identity, Social; Postmodernism

Further Readings

Assmann, J. (2011). *Cultural memory and early civilization: Writing, remembrance, and political imagination.* Cambridge, England: Cambridge University Press.

Halbwachs, M. (1992). *On collective memory.* Chicago, IL: University of Chicago Press.

Hutton, P. (1993). *History as an art of memory.* Burlington, VT: University of Vermont Press.

Levy, D., & Sznaider, N. (2006). *The holocaust and memory in the global age.* Philadelphia, PA: Temple University Press.

Nora, P. (1996–1997). *Realms of memory: Rethinking the French past.* New York, NY: Columbia University Press.

Olick, J. K., Vinitzky-Seroussi, V., & Levy, D. (Eds.). (2011). *The collective memory reader.* Oxford, England: Oxford University Press.

Young, J. (1993). *The texture of memory: Holocaust memorials and meaning.* New Haven, CT: Yale University Press.

Collective Moral Responsibility

This entry discusses an important feature of collective entities and social life in general, that of moral responsibility ascribed to collective agents or plural subjects. It first explains the distinctive feature of collective moral responsibility and then goes on to examine briefly three main theories about it.

What Is Collective Moral Responsibility?

Collective moral responsibility is a species of moral responsibility. Here, we need to distinguish *moral* responsibility (including collective moral responsibility) from *causal* responsibility. A person or persons can inadvertently cause a bad outcome without necessarily being morally responsible for doing so. For example, a careful and competent driver who is obeying all the rules of the road might, nevertheless, accidently and unavoidably hit and maim a child who suddenly and unpredictably runs onto the road and in front of the moving vehicle. Moral responsibility typically requires not only causal responsibility but also an *intention* to cause harm (or at least the knowledge that one's action will or may well cause harm).

Collective moral responsibility is the moral responsibility that attaches to *structured* and *unstructured groups* for their morally significant actions and omissions. Thus, an organized gang of thieves who carry out a million-dollar bank heist is said to be collectively morally (and, presumably, legally) responsible for the theft. Again, a number of bystanders who act jointly to save a child trapped in a burning house are said to be collectively morally responsible for saving the child's life. Notice that sometimes it is the *members* of a group that are said to be collectively responsible and sometimes it is *the group or other collective entity per se*; for example, the Mafia might be said to be collectively morally responsible for a crime wave in southern Italy.

Theories of Collective Moral Responsibility

There are three prominent kinds of theories of collective moral responsibility. The first of these conceives of collective moral responsibility as a convenient way of referring to what is in fact simply a set of individual responsibilities. I will refer to it as the atomistic account. The second holds that it is the group or collective itself that is the bearer of moral responsibility. I will refer to this view as the collectivist account. The third theory is a relational account. The only bearers of moral responsibility are individual human persons (or like creatures) but, unlike individual moral responsibility, collective moral responsibility is to be understood in relational terms.

Atomism

Advocates of atomism with respect to collective moral responsibility include H. D. Lewis, R. S. Downie, Stephen Sverdlik, and Jan Narveson. The strength of atomism is that it does not postulate suprahuman collective entities as the mysterious bearers of moral responsibility and of the psychological states (e.g., beliefs and intentions) necessary for moral responsibility. Moreover, since it ascribes moral responsibility only to individual human beings, it has no tendency to let the members of collective entities, such as criminal organizations or negligent corporations, off the hook by relocating moral responsibility at the suprahuman level.

The weakness of atomism is that it does not seem to be able to accommodate the full range of cases in which we pretheoretically ascribe collective moral responsibility. For example, in our above million-dollar bank heist example, robber A might be individually responsible for planning the heist, robber B for driving the getaway car, C for blowing the safe, D for taking $100,000, E for taking a second $100,000, and so on. However, arguably none of the robbers was individually causally responsible—and, therefore, individually morally responsible—for stealing $1,000,000. Again, we pretheoretically hold BP morally responsible for the massively environmentally damaging oil spill in the Gulf of Mexico in 2009. But here we seem to be doing something above and beyond simply ascribing individual moral responsibility to each of the BP personnel who did something wrong and aggregating them. BP's moral responsibility seems to be more than the set of individual responsibilities.

Collectivism

In this view, "collective responsibility" should be understood in the sense of a collective's responsibility. Whether the members of a collective, the people

who constitute the collective in question, are individually responsible is a separate question. When it comes to collective moral responsibility, the collective itself is the bearer of the moral responsibility. In David Copp's terminology, this claim is equivalent to the claim that a collective can be an independent moral agent. An important corollary of the collectivist view is that collectives are capable of bearing moral responsibility for outcomes, even when none of their members are in any degree *individually* morally responsible for those outcomes. Other prominent representatives of the collectivist approach include Peter French and Margaret Gilbert.

Perhaps unsurprisingly, collectivism tends to be regarded as being able to accommodate many of the problems that beset atomism but to lack the virtues of atomism. Thus, to return to our examples, collectivism will ascribe moral responsibility to (respectively) the gang of robbers as a collective entity and to BP per se (and not merely to some individual BP personnel for their individual moral failings). However, in so doing it admits suprahuman collective entities (the gang, BP) that (somewhat mysteriously) bear moral responsibility (and, therefore, the associated psychological states) and have the potential to get (respectively) the individual gang members and BP personnel off the moral hook. Potentially at least, BP is morally responsible, but none of its managers, employees, and so on have any moral responsibility; and the same goes for the individual gang members, at least as far as the theft of the $1,000,000 is concerned (as opposed to, say, driving the getaway car or stealing $100,000).

Relational Accounts

Relational accounts try to square the circle. Theorists of a relational persuasion include Larry May, Gregory Mellema, and Seumas Miller. On the one hand, relational accounts claim to have the virtues of atomism and to lack the vices of collectivism. Thus, relational accounts do not postulate suprahuman collective entities as the mysterious bearers of moral responsibility and have no tendency to let the members of collective entities off the moral hook by relocating moral responsibility at the suprahuman level.

On the other hand, relational accounts seem to be better equipped than atomistic accounts to provide a plausible account of the ascription of moral responsibility to structured and unstructured groups.

For example, Miller has developed an account of collective responsibility as joint responsibility. On Miller's view, at least one of the central senses of collective responsibility is responsibility arising from joint actions and omissions. Roughly speaking, a joint action can be understood thus: Two or more individuals perform a joint action if each of them intentionally performs an individual action but does so with the (true) belief that in so doing they will jointly realize an end that each of them has. On this view of collective responsibility as joint responsibility, collective responsibility is ascribed to individuals. Each member of the group is individually morally responsible for the outcome of the joint action. However, each is individually responsible, jointly with the others; hence, the conception is relational in character. Thus, in our million-dollar bank heist example, each member of the gang is responsible jointly with the others for the theft of the $1,000,000, because each performed his contributory action in the service of that collective end (the theft of the $1,000,000).

On this kind of relational view, BP can be ascribed collective moral responsibility for the Gulf oil spill to the extent that BP personnel jointly acted—or, more likely, jointly (and culpably) failed to act—in ways that led to the disaster. Here, the network of joint actions and omissions could be quite wide and complex without involving all, or even most, BP personnel. Moreover, some joint actions or omissions might be of greater moral significance than others and some individual contributions, such as those of managers, of greater importance than others.

Seumas Miller

See also Collective Agents; Collective Goals; Collective Values; Group Mind; Holism, in the Social Sciences; Individualism, Methodological; Institutions as Moral Persons

Further Readings

Copp, D. (2006). On the agency of certain collective entities. *Midwest Studies in Philosophy, 30,* 194–220.

Downie, R. S. (1969). Collective responsibility. *Philosophy, 44,* 66–69.

French, P. (1984). *Collective and corporate responsibility.* New York, NY: Columbia University Press.

Gilbert, M. (2002). Collective guilt and collective guilt feelings. *Journal of Ethics, 6*(2), 115–143.

Lewis, H. D. (1948). Collective responsibility. *Philosophy, 24,* 3–18.
May, L. (1992). *Sharing responsibility.* Chicago, IL: University of Chicago Press.
Mellema, G. (1988). *Individuals, groups and shared moral responsibility.* New York, NY: Peter Lang.
Miller, S. (2006). Collective moral responsibility: An individualist account. *Midwest Studies in Philosophy, 30,* 176–193.
Narveson, J. (2002). Collective responsibility. *Journal of Ethics, 6,* 179–198.
Sverdlik, S. (1987). Collective responsibility. *Philosophical Studies, 51,* 61–76.
Zimmerman, M. J. (1985). Sharing responsibility. *American Philosophical Quarterly, 22*(22), 115–122.

Collective Rationality

Collective rationality is a property of a group of agents. In one sense, it is the rationality of each member of the group, but in another, common sense, it is the rationality of the group itself.

This entry treats the rationality of the group itself but also describes its relation to the rationality of each member of the group. Being normative, collective rationality falls into philosophy's province. Because groups often aim for the collectively rational, collective rationality figures in the explanation of their acts and so assists the social sciences. The entry's first section explains collective rationality, and the other sections treat its application to collective acts.

Agents and Rationality

Groups act; a group's acts may be rational or irrational. A group may pass a resolution to raise dues. Its act may be rational because the group needs more revenue to balance its budget. A group may adopt a policy of not accepting new members. The policy's adoption may be irrational because the group needs new members to survive.

As for individuals, the standards of rationality for a group are sensitive to the group's situation. Suppose that the members of a group vote rationally, although they do not vote in the best way because of their ignorance. The group then votes rationally, although not optimally. Its members' ignorance may excuse the group's shortcoming.

Collective rationality is a technical term in some literatures. In economics, it often means efficiency, or Pareto optimality (after the Italian economist Vilfredo Pareto), and in some contexts, it means maximization of collective utility, taken as a sum of utilities for individuals. This section explains each of these technical senses of collective rationality, starting with Pareto optimality.

Suppose that a group has two options and the first is better than the second for each member of the group. Then, the first option is strictly Pareto superior to the second option. A group's act is weakly Pareto optimal if no alternative is strictly Pareto superior to it. Suppose that a group has two options and the first is better than the second in some cases and at least as good as the second in all cases. Then, the first option is Pareto superior to the second option. A group's act is Pareto optimal if no alternative is Pareto superior to it. A group benefits from achieving Pareto optimality. However, in some cases a group composed of rational people may fail to achieve Pareto optimality. The members of the group, because of nonculpable ignorance, may not know which of their options would be Pareto optimal.

An act's utility for a group is the sum of the personal utilities that the group's members attach to the act. Their personal utilities depend on how strongly they want the group to perform the act. An act that maximizes collective utility is such that no other act achieves greater collective utility. Typically, giving each member of a group a resource has greater collective utility than withholding the resource from each member of the group.

A Pareto-optimal act may not maximize collective utility. For example, in a typical case, every way of dividing a sum of money among a group's members achieves Pareto optimality. Every alternative gives less money to some member, and so none is Pareto superior. Typically, not all divisions maximize collective utility, however. For example, giving equal shares to the group's members yields more collective utility than giving all the money to the flush members and nothing to the needy members.

Achieving Pareto optimality seems insufficient for collective rationality, so instead of taking collective rationality to be Pareto optimality, some theorists take it to be maximization of collective utility. Being sensitive to a group's circumstances, rationality excuses some failures to maximize collective utility. A group composed of rational people, through nonculpable ignorance, may fail to

maximize collective utility because they fail to see which of their options maximize collective utility. Their failure to maximize collective utility may not be collectively irrational, all things considered.

The standard of collective-utility maximization requires a way of putting personal-utility assignments on the same interpersonal scale before summing them to obtain collective utility. Some theorists object to interpersonal utilities because it may be difficult to measure them accurately. The standard of Pareto optimality is easier to apply than is the standard of collective-utility maximization.

Standards of collective rationality apply to the traits of a group, such as a group's customs and traditions. Some customs may be rational and others irrational. Standards also apply to a group's acts, such as electing a president. A group's act is often called a collective act.

What counts as a collective act? Because ordinary usage does not settle the question, theorists are free to define collective acts in any fruitful way. Some hold that a collective act requires a collective intention in a sense that does not require a group to have a mind. They hold that a person's having an intention to perform an act is a functional state of the person that normally generates the act. In this functional sense, whatever plays the role of an intention in generating a group's act counts as a collective intention. In a restrictive sense, only those acts a group performs because of a collective intention count as collective acts.

A liberal view of collective acts counts all sets of acts of a group's members, with one act per member, as collective acts. If Arthur breaks a pencil in England and Luis breaks a pencil in Uruguay, then between them they break two pencils. Neither breaks two pencils, but the two of them break two pencils. Together they break two pencils, although they did not coordinate to break them. Given the broad view of collective acts, not all of a group's collective acts count as the members coordinating, cooperating, or collaborating. The broad view acknowledges special types of collective acts, such as coordination, that require a group's members to take account of each other. Some, but not all, collective acts achieve coordination.

To make standards of collective rationality as far-reaching as possible, a broad theory of collective rationality adopts the liberal view of collective acts. It evaluates for rationality even unconnected acts of a group's members.

This entry treats the rationality of collective acts and puts aside evaluation of a group's traits. What standards of rationality apply to collective acts? Some standards for individuals suggest extensions to groups. One suggestion holds that a rational collective act comes from the top of a collective preference ranking of acts. This principle requires a definition of collective preference, however, and no satisfactory general definition using only the preferences of the group's members exists, as Kenneth Arrow shows.

Utilitarianism offers another way of aggregating personal assessments of options to form a collective assessment. It suggests that as a person's act maximizes personal utility if it is rational, a group's act maximizes collective utility if it is rational, at least if conditions are ideal for collective action. If contrary to the principle's presupposition, interpersonal utilities for a group's members do not exist in all cases, the principle may restrict itself to cases in which the interpersonal utilities exist.

The principle to maximize collective utility faces a serious problem. It may require sacrificing a member of a group. This happens if the group can prevent the deaths of many members its enemy holds hostage by exchanging for the hostages one hunted member its enemy will execute. The member then rationally blocks maximization if possible. The principle pits collective rationality against individual rationality. This makes the theory of rationality inconsistent. The theory tells a group to maximize collective utility but tells a member to block maximization. The theory should issue consistent directives to the group and to its members.

An alternative principle for evaluating collective acts holds that the rationality of every member of a group ensures the rationality of the whole group. If each member acts rationally, then the collective act that their acts form is also rational. For example, if all the members vote rationally, then the group's vote is also rational. This principle of evaluation is general and ensures consistency. According to it, rationality's directives to a group are not contrary to its directives to the group's members.

Granting the principle of evaluation, the rationality of all members of a group suffices for the group's rationality. Is the rationality of all members also necessary for the group's rationality? No, in some cases a group's act is rational, although some

members' contributions are irrational. For example, a committee using majoritarian methods to vote on resolutions may rationally vote in favor of a sensible resolution, even though some members irrationally vote against it. Because a rational act by the group requires only a majority of rational votes, a few irrational votes do not make the group's act irrational. What makes the group's voting rational despite the irrational voting of some members? The group's act is rational because it is the same collective act that the group would perform if all its members were to vote rationally.

A group's profile of votes is a description of every member's vote. Even a few irrational votes make a group's profile of votes defective. In the example, the group's profile of votes is irrational, one says, although its voting for the resolution is rational. To accommodate this point, a theory of rationality may distinguish the group's profile of votes from the group's voting for the resolution. Although the group's profile of votes constitutes the group's voting for the resolution, the collective acts differ because they are individuated according to the propositions that express them. Two acts that have the same realization by physical events may differ because they have different propositional expressions. One may be rational, while the other is irrational. A group's profile of votes may be irrational, although the group's voting for the resolution is rational.

The Prisoner's Dilemma

In the Prisoner's Dilemma, each of two agents, who cannot communicate, decides whether to act cooperatively. If both act cooperatively, they each do better than if each acts noncooperatively. However, each does better acting noncooperatively whatever the other agent does. Each agent prefers the outcome in a pair that is better for him. A principle of rational preference, the principle of strict dominance says that if one option is better than another in every case, then the agent should prefer it to the other option. The principle requires that each agent act noncooperatively even though the combination of their noncooperative acts is not Pareto optimal. It is not Pareto optimal because each agent prefers the combination of his or her cooperative acts. Because some theorists take collective rationality to require Pareto optimality, they conclude that the principles of individual rationality conflict with the principles of collective rationality. The principle of dominance and the principle of Pareto optimality are at odds in the Prisoner's Dilemma. The first requires that each agent act noncooperatively, whereas the second requires that the pair of players act cooperatively to achieve a Pareto-optimal combination.

Collective and individual rationality do not conflict, contrary to the foregoing account of their application to the Prisoner's Dilemma. Collective rationality does not always require Pareto optimality. Sometimes obstacles stand in the way, obstacles such as the inability to communicate. In the Prisoner's Dilemma, agents cannot communicate. Conditions are not ideal for collective action. Hence, collective rationality does not demand that the agents achieve Pareto optimality. Collective rationality, as individual rationality, is sensitive to circumstances. It excuses failures to optimize given the obstacles to optimization.

Group Minds

Groups literally act but lack minds and so do not literally have beliefs and desires, or intentions, and do not literally make decisions. A technical definition of collective preference may be helpful, nonetheless, in special cases where restrictions prevent problems. Take a definition of collective preference according to which a group prefers the first of two options if a majority of the group's members prefer the first option. This definition leads to cyclical collective preferences if advanced generally. To prevent cycles, the definition may restrict itself to cases in which the group's members have preferences that are single-peaked. This restriction requires that according to some ordering of alternatives along the horizontal axis of a graph, a plot of each individual's preference ranking of alternatives, with height representing position in the ranking, has a single peak. With the restriction in place, the majoritarian definition of collective preference works well. An account of the group's behavior may treat the group as if it had collective preferences.

Game Theory

In a game of strategy, each player chooses a strategy, and the combination of strategies chosen settles the outcome for each player. Strategic reasoning suggests that a player choose a strategy in light of his beliefs about the strategies that other players will choose. A profile of strategies is a set of strategies with exactly one strategy for each player. A solution

to a game is a profile of strategies such that each strategy is rational given the other strategies in the profile. Realization of a profile of strategies is a collective act. How are the profile's being a solution and its being collectively rational related?

In ideal games, agents are fully rational and have unlimited cognitive power so that they reason flawlessly and know all the logical and mathematical truths. They also have correct beliefs about others' strategies. These beliefs may arise from their common knowledge of their game and their rationality. Here, common knowledge has a technical sense according to which common knowledge that a proposition holds entails that all know the proposition, all know that all know it, and so on.

Because in ideal games each player has correct beliefs about the other players' strategies, if each player's strategy maximizes utility, the players' strategy profile is what is known as a Nash equilibrium (after John Nash). Each player's strategy is a best response to the other players' strategies. A player's strategies in a game may include mixed strategies, that is, probabilistic mixtures of the player's pure strategies. Every game with a finite number of players and pure strategies has a Nash equilibrium if the players may choose mixed strategies. Game theorists propose that a strategy profile is not a solution unless it forms a Nash equilibrium, assuming that the game has a Nash equilibrium. Some games have multiple Nash equilibria. Theorists propose principles for selecting the Nash equilibria that qualify as solutions. A common requirement is that a solution be a Nash equilibrium that is Pareto optimal among Nash equilibria.

Because a game's solution is a strategy profile in which each player's strategy is rational given the other players' strategies and because the rationality of a player's strategy depends on the player's information, the strategy profiles that are solutions depend on the players' information about their game and each other. An adequate representation of a game specifies the players' information.

The definition of a solution uses conditional as opposed to unconditional rationality. Each player's part of a solution is rational if each player knows that the solution will be realized. This result need not hold for a profile of rational strategies. A player's strategy in the profile may be rational only because the player thinks that other players will perform strategies outside the profile. For example, in the game of Matching Pennies, the player going for a match may rationally show heads because he thinks his opponent will show heads. The player going for a mismatch may show tails because she thinks her opponent will show heads. The two strategies are each rational in this case, although they are not each rational if both players know about them. Given that knowledge, the player going for a match should not show heads.

Given that the rationality of each player's strategy ensures the collective rationality of the players' strategy profile, a collectively rational profile may fail to be a solution. In the example, the matcher's showing heads and the mismatcher's showing tails is a profile of rational strategies and so is collectively rational, but it is not a solution. However, if a strategy profile is a solution, then given the players' awareness of their participation in the profile, each player's participation is rational and the player's realization of the profile is collectively rational.

Many theorists argue that in some games the players' rationality does not ensure a solution. In the game Hi-Lo, the players have two ways of coordinating, one better for each player than the other. Each has a reason to pick High, his part in the better way of coordinating, if the other player picks High. If neither player is told what the other will do, it may seem that none has a reason to pick High. Nonetheless, each player's picking High appears to be the game's solution.

Must collective rationality demand more than the rationality of each player if it is to generate a solution? If the game has fully rational ideal players who are prepared for coordination, an argument exists that each player chooses High. One player may intend to pick High because forming the intention gives the other player a reason to pick High. The first player then has a reason to pick High because the second player will. The players' rationality in preparing for optimal forms of coordination leads to the profile in which each player picks High.

By describing solutions to ideal games, theorists flesh out an account of collectively rational behavior. Because players infer the profile realized, realizing a solution ensures the players' collective rationality.

Paul Weirich

See also Collective Agents; Collective Intentionality; Decision Theory; Game-Theoretic Modeling; Judgment Aggregation and the Discursive Dilemma;

Pareto Optimality; Rational Choice and Political Science; Rationality and Social Explanation; Social Choice Theory; Team Reasoning

Further Readings

Arrow, K. (1951). *Social choice and individual values*. New Haven, CT: Yale University Press.

Bacharach, M. (2006). *Beyond individual choice: Teams and frames in game theory* (N. Gold & R. Sugden, Eds.). Princeton, NJ: Princeton University Press.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge, England: Cambridge University Press.

Bratman, M. (1992). Shared cooperative activity. *Philosophical Review, 101*(2), 327–341.

Copp, D. (1995). *Morality, normativity, and society*. New York, NY: Oxford University Press.

Hardin, R. (1982). *Collective action*. Baltimore, MD: Johns Hopkins University Press.

List, C., & Pettit, P. (2011). *Group agency: The possibility, design and status of corporate agents*. Oxford, England: Oxford University Press.

Ludwig, K. (2007). Collective intentional behavior from the standpoint of semantics. *Noûs, 41*, 355–393.

McMahon, C. (2001). *Collective rationality and collective reasoning*. Cambridge, England: Cambridge University Press.

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge, England: Cambridge University Press.

Weirich, P. (2010). *Collective rationality: Equilibrium in cooperative games*. New York, NY: Oxford University Press.

Collective Values

This entry gives an account of the epistemological and ontological issues behind the notion of collective values, explains its difference from that of shared values, and shows the importance of admitting collective values in social ontology.

On March 16, 2011, Rajav Shah, the head of the U.S. Agency for International Development (USAID), presented his budget proposal to the House Subcommittee on Foreign Affairs. He argued that foreign assistance from USAID "keeps our country safe, develops the markets of tomorrow, and expresses our collective values." While appeals to our *collective values* are often politically expedient, the content of such appeals is often ambiguous at best. Of course, collective values must be *values,* not mere common expectations or shared understandings of joint activities. They require treating activities, entities, or practices as worthwhile or essential to what we do together. Furthermore, collective values are not merely agent-neutral values; they must be grounded in our collective concern with activities such as the elimination of sex trafficking, the prevention of deforestation, and the promotion of education.

But while the authors of this entry value such things, and one suspects that most other U.S. citizens do as well, this does not mean that they are *American values*—values that are not just mine and yours, his and hers, but *ours*.

If the methodologically and ontologically individualistic assumption that *there are no collectivities* is true, then there are no collective values; rather, appeals to collective values are literally false, though often elliptical for true claims about aggregations of individual values. Perhaps the claim "Justice is an American value" might be better paraphrased as "Every U.S. citizen values justice" or as "Most U.S. citizens value justice."

Unfortunately, such paraphrases threaten to flatten an important distinction between *shared* and *collective* values, which can be exemplified in the difference between the members of an academic department who all happen to value good baseball games, and the department's valuing of a specific deliberative procedure. The baseball value could be kept private; but even when it is made public it does not have an impact on department-relevant behavior. The valuing of a deliberative procedure, a *collective* value, by contrast, grounds the practices and projects of the department *as such*. So while "shared values" and "collective values" are often treated as interchangeable, they are logically distinct.

Collective and shared values are significant for understanding some collectivities. For example, someone who does not understand their near-religious exaltation of cheese curds and Usinger's sausage may not really understand Wisconsinites. Moreover, shared and collective values are significant for decision making, as when marketing to a particular demographic requires considering their shared values. However, the fact that people *happen*

to share a value is unlikely to affect action in any way that extends beyond the action guidance of individual values.

Of course, a more sophisticated form of methodological individualism might couple the individualistic requirement on explanation with a nonindividualistic conception of *value* (e.g., as Max Weber famously did). The seeming paradox of this position dissipates with the recognition that networks of social practice can be seen as the structural scaffolding upon which individual values can be constructed or as the holistic conceptual background against which individual practices of valuing can emerge. For example, valuing the vintage New England Patriots football logo requires the existence of numerous social institutions (e.g., facts about national and local history, the existence of the National Football League, and fashion trends at the time of the American Revolution). But such a value is not a *collective* value; while social institutions are necessary to explain how such values are possible, the individual is still the locus of valuation and the locus of action.

In short, genuinely *collective values* require (a) a collective agent capable of goal-directed behavior and (b) collective values that are implicated in, required for, or constitutive of that behavior. This does not require an ascription of value *to the collectivity,* and some collective values are likely to be dependent on, though not reducible to, individual values. Such values arise where the social institutions that are a necessary condition for individual valuing are grounded in the collective nature of a joint activity. Call these "we-values" (to parallel "we-intentions"). Unlike values contingently shared by group members, we-values depend on a complex set of relationships that can modulate the ways in which individuals reflect upon these values, consider conflicting values, and recalibrate their individual and collective actions to accord with those values. Some we-values are analytically constitutive of the group itself (e.g., the Society for the Preservation of Greek Heritage must treat Greek culture as worthy of value), while other we-values achieve their status as a result of the necessity of the value for continuing a collective endeavor. For example, it is an essential part of belonging to the modern scientific community that one values seeking the truth, replicating results, and preserving methodological transparency.

These we-values may be relatively thick or relatively thin. A rabbit values her life because all of her actions accord with this value, and such relatively thin values thus offer a way of categorizing and describing her behavior. Similarly, we might say that a particular corporation values an increase in its profits because its actions accord with this value. Such "free-floating values," however, are insufficient to explain the role of valuing in individual and collective deliberation. Agency in a thicker sense may well require reflecting on one's values, considering how they interact and conflict, and attempting to bring one's actions into accord with one's values. Thick values, unlike their thin counterparts, provide structure for deliberations and play an integral role in practical reasoning.

This brings us to the most *robust* collective values, values that are properly predicated *of the collectivity itself* (as Margaret Gilbert claims). In such cases, *the collectivity* must be able to reflect on its values, consider how they interact and conflict, and evaluate the extent to which individual actions should be brought into accordance with these values. For example, a university that has plans and projects designed to ensure that its actions (e.g., admissions and hiring decisions) increase diversity, and endeavors to alter its actions where they do not align with this value, can be properly said to value diversity.

Collective values provide action guidance that stretches well beyond the recognition of aggregates of individual values. They change the deontic status of particular actions for group members. Although a new member of an academic department, for example, has no more reason to value fine wine after learning that her colleagues do so than she does when she learns that some loose aggregate of her friends do, when she learns that her colleagues value improving gender equality in her profession, she gains new reasons for action (and there is a chance that—as a member of this group—she will have to give up some reasons that she used to have). Of course, the fact that something is valued by a collectivity to which one belongs does not imply that such values ought not be changed or challenged, but as a group member such challenges are the exception, not the rule. Perhaps less obviously, collective values also can change the deontic significance of actions for nonmembers, making actions more egregiously wrong than they would be if merely shared by a random set of agents. For example, there is an

additional reason to respect an artifact that is valued by a group for its role in their collective activities, history, or self-understanding that goes beyond the reason to respect a similar artifact that is valued by a similar number of disconnected individuals. So the fact that a value is collective rather than merely shared can be morally significant both for members and for those external to a collectivity.

Bryce Huebner and Marcus Hedahl

See also Collective Agents; Collective Emotions; Collective Goals; Collective Intentionality; Collective Rationality; Group Beliefs; Group Identity; Group Mind; Holism, in the Social Sciences; Individualism, Methodological; Social Norms; Social Ontology, Recent Theories of

Further Readings

Anderson, E. (2004). *Value in ethics and economics*. Cambridge, MA: Harvard University Press.

Gilbert, M. (2000). *Sociality and responsibility*. Lanham, MD: Rowman & Littlefield.

Graham, K. (2002). *Practical reasoning in a social world*. Cambridge, England: Cambridge University Press.

Kitcher, P. (2001). *Science, truth, and democracy*. Oxford, England: Oxford University Press.

Sistare, C., May, L., & Francis, L. (Eds.). (2001). *Groups and group rights*. Lawrence: University of Kansas Press.

Weber, M. (1978). *Economy and society* (G. Roth & C. Wittich, Eds.). Berkeley: University of California Press. (Original work published 1922 in German as *Wirtschaft und Gesellschaft*, Tubingen, Germany: Mohr)

Commitment

This entry explains the notion of commitment and its significance in philosophical accounts of social agency and presents the three received accounts of it.

The Notion of "Commitment"

The English word *commitment* has many meanings and is without equivalent in other languages. In the sense that is particularly relevant to philosophy and the social sciences, the term refers to an agent's choice or disposition to self-impose limitations on her agency ("self-commitment"). While changes of options and elimination of alternatives occur as a side effect or unintended consequence of almost any choice, it is usually assumed that in the special case of commitment, the agent *identifies* with these self-imposed bounds of her agency as expressive of her will. Commitments are a central feature of human agency and a basic building block of the social world.

The Received Accounts of Commitment

Received accounts of the nature and role of commitment can be distinguished according to how far they depart from the standard rational choice model of action (the standard model assumes that an agent's choices maximize the satisfaction of her desires, given her beliefs). Broadly, the received conceptions of commitment can be divided into three groups: (1) according to the *conservative view,* commitment leaves the standard model more or less intact; (2) proponents of *revisionist accounts* claim that the view of practical reason has to be widened in order to accommodate commitment; and (3) *revolutionary conceptions* claim that in view of the structure and role of commitment in action, an altogether different theory of practical reason is required.

Conservative Accounts

Diverse theorists such as Thomas C. Schelling or Jon Elster analyze a classical paradigm case of commitment: the capacity to remain steadfast in the face of temptation. In their view, this is explained by the human capacity to causally influence the range and utility of future options, as illustrated in Odysseus's case of having himself tied to the mast of his ship so as not to be able to give in to temptation. In political theory, the role of constitutional constraints imposed on collective action is sometimes analyzed along similar lines. Commitments are seen as a special case of rational choice, as the source of the commitment is placed in the preferences of the user of the "commitment device" rather than in the exertion of a different dimension of evaluation.

Revisionist Accounts

According to revisionist authors, the kind of self-control involved in commitment requires the exercise of a special type of mental capacity or practical reason rather than the merely causal power that an agent has over his own future self. Commitment does not *remove* options from the menu of available

alternatives but involves a different, noninstrumental type of *evaluation* according to which choices are made. In these accounts, "commitment" refers to an agent's ability or disposition to be bound—and guided in the course of her actions—by what she takes herself to have a *reason* to do, independently of other motivations he or she might have. While the Kantian allusions are certainly not coincidental, it is important to note that according to most revisionist accounts, an agent's commitments need not express her acceptance of universal moral or rational principles but may stem from the particular loyalties, values, or other normative bounds (or plans) that she endorses either as a matter of an autonomous act of will or as an effect of socialization. Some variants of these conceptions concentrate on the "deep," temporally stable normative bounds that are claimed to be constitutive of an agent's personhood, authenticity, or integrity (e.g., "identity-conferring commitments"), while other accounts focus on ubiquitous kinds of commitments involved in any case of rule following or acting in accordance to a plan. Margaret Gilbert has developed an account of joint commitment that she claims to be constitutive of social facts. A joint commitment to x is created if the participant individuals implicitly or explicitly express the willingness to participate in the group's joint x-ing, and it puts the members under an obligation to perform their parts, independently of changes in their motivation. A joint commitment may not be unilaterally rescinded.

Revolutionary Accounts

At the revolutionary extreme of the spectrum are those philosophers (known under labels such as "interpretivists," "holists," or "social externalists") who hold that for any intentional (mental) state to have content (e.g., for "choice" to be "of" something in the sense that the agent has any grasp *of what* it is that she chooses), the agent needs to be committed to a rule. Thus, commitment is seen as a basic and irreducible feature of mind and action. Most prominently in the philosophy of social science, Peter Winch claims that in order to understand an agent's behavior as an action the interpreter must understand the agent as committed to the rules that constitute the meaning of the action in question. Along similar lines, proponents of inferentialism and normative pragmatism, such as Robert Brandom, argue for a view of practical reason in which endorsement and commitment are the central features of rational agency, rather than notions such as desire or preference, traditionally understood as the referents of the "springs of action."

An account of commitment that is especially important for the social sciences can be found in Amartya Sen's critique of standard rational choice theory. Sen claims that committed action is pervasive in social life (e.g., linguistic communication, cooperation, adherence to moral codes and loyalties) and that it cannot be accommodated in standard rational choice explanations. Sen distinguishes commitment from *sympathy.* A sympathetic agent furthers another's welfare because it correlates with his own. A committed agent, by contrast, may act in other people's interests independently of any such effect on his or her welfare. Commitment thus drives a wedge between choice and welfare and involves the possibility of "counterpreferential choice." In this context, Sen advanced the radical claim that committed agents should not be construed as pursuing their own goals. While some interpreters have linked this idea to the role of social identities and shared goals, Sen has later explained his claim as referring to an agent's conforming to a norm independently of how the agent takes this to affect his own or other agents' welfare.

Hans Bernhard Schmid

See also Decision Theory; Rational Expectations; Rationality and Social Explanation; Rule Following; Sen's Paretian Liberal; Social Norms

Further Readings

Brandom, R. B. (2000). *Articulating reasons.* Cambridge, MA: Harvard University Press.

Elster, J. (1979). *Ulysses and the sirens.* Cambridge, England: Cambridge University Press.

Peter, F., & Schmid, H. B. (Eds.). (2007). *Rationality and commitment.* Oxford, England: Oxford University Press.

Peter, F., & Spiekermann, K. (2011). Rules, norms and commitments. In I. C. Jarvie & J. Zamora-Bonilla (Eds.), *The Sage handbook of the philosophy of social sciences* (pp. 216–238). Thousand Oaks, CA: Sage.

Sen, A. K. (1985). Goals, commitments, and identity. *Journal of Law, Economics, and Organization,* 1/2, 341–355.

Winch, P. (1958). *The idea of a social science and its relation to philosophy.* London, England: Routledge.

Common Goods

This entry considers an important category of collective entities known as *common goods* and reviews the different notions, distinctions, and problems related to such a collective item. The entry raises questions about what it means to be "common" in this case and how it relates to the issues of jointly held rights to such common goods, which are jointly produced or are the result of joint activity.

Common (or collective) goods have been theorized by philosophy, by political theory, by jurisprudence, as well as by social sciences, such as economics.

There is evidently a family resemblance between notions such as *common good, collective good, public good, common interest, collective interest, public interest,* and so on. Such goods or interests attach to or are enjoyed by groups or other collectives, such as, for example, the British people, the members of the Wagga Wagga local community, the teachers' union, or the pharmaceutical industry. So the contrast here is with a single individual's interest or a benefit that is or could be produced and/or enjoyed by a single individual. Historically, notions of the common good, common interest, and the like in the political sphere are associated with philosophers such as Aristotle, Thomas Aquinas, Thomas Hobbes, and Jean-Jacques Rousseau.

Notions of Common Goods

There is a distinction to be made between the common good and specific common goods. Perhaps security, clean air, and an efficient transport system are all examples of common goods. We can presumably, at least in principle, offer a definition of the notion of a common good and draw up a list of such goods. By contrast, *the common good*—which is often, but not necessarily always, what is in the common interest—is something to be determined anew in a multiplicity or ever-changing circumstances. The common good is an unspecified, or at least underdetermined, state to be realized by collective action.

Economists typically speak of a species of common goods, namely, so-called public goods. They define public goods as being nonrival and nonexcludable. If a good is nonrival, then my enjoyment of it does not prevent or diminish the possibility of your enjoyment of it; a street sign, for example, is nonrival since my using it to find my way has no effect on your using it likewise. Again, a good is nonexcludable if it is such that if anyone is enjoying the good then no one can be prevented from enjoying it—for example, national defense. The public goods in question are typically relativized to the nation-state but increasingly to the global economy.

Nonrivalness and nonexcludability are relevant to the characterization of common goods, albeit the notion of a common good is not necessarily defined in terms of them. Other properties that are relevant to the notion of a common good include equality and jointness of production. Many common goods are jointly produced (or maintained or renewed). And perhaps if a common good is enjoyed, it is enjoyed equally by all; if not, it should be.

There are further distinctions to be made in relation to common goods (sometimes referred to as *collective goods*).

As Joseph Raz points out, there are *necessarily* common goods and ones that are merely *contingently* common. A right of access to a water supply might only be contingently common. This would be so if when the water supply is cut off, everybody's supply is cut off. But under a different system, selective cutting off is possible. By contrast, a tolerant society is necessarily a common good. The tolerance of the society is not something that could be channelled to certain individuals only.

Following Denise Reaume, in respect of necessarily common goods, we can further distinguish between those that an agent can choose not to enjoy himself and those that he cannot choose not to enjoy. Perhaps clean streets are of the former kind and a law-abiding society of the latter. A recluse could not be prevented from enjoying clean streets or a law-abiding society, but he could choose not to enjoy clean streets by never going out. On the other hand, even by staying at home he cannot choose not to enjoy a law-abiding society.

Jointly Produced Common Goods and the Question of Rights

Let us now further explore common goods that are *jointly produced* (or maintained or renewed). Perhaps the territory occupied by the people of the Netherlands is a common good in this sense, since

much of it would be under water were it not for the elaborate system of dykes put in place, maintained, and extended over hundreds of years by the people of the Netherlands.

What is the relationship, if any, between moral rights and jointly produced common goods? Presumably the participants in the joint enterprise in question have a joint moral right to the common good. Accordingly, even if the good considered in itself is not a common good, each of the individual (jointly held) rights to it is a right to it qua common good. For example, by using two boats and a single large net we could jointly catch 100 fish. By prior agreement, we could possess individual rights to 50 fish each. But this agreement is something additional to the joint right, and indeed presupposes it. Imagine that unexpectedly the good produced could not be parcelled out in the manner envisaged in the agreement. Perhaps we caught only one very large fish or no fish but instead a rare and valuable old ship. If so, each individual could still claim an individual (jointly held) right to the good and therefore legitimately insist on making some different agreement (or perhaps no agreement).

Jointly produced common goods give rise to the so-called free rider problem. The problem arises for goods that can be produced even if some members of the group do not contribute; that is, they free ride. The problem is that each knows he or she can free ride and it is in his or her interest to do so. However, if all act on their self-interest, then the good will not be produced, something that is ultimately in no one's interest. The problem is particularly acute for rational egoist theories (e.g., Hobbes's classic one), since such theories assume that rational human action is always self-interested or, at least, that where self-interest and the common good conflict, self-interest always wins.

Common Goods Produced by Joint Activity

Thus far, our discussion of jointly held rights to jointly produced goods has implicitly assumed that the good produced is not constituted by the actions that produce it. Let us now consider goods that consist of the *joint activity* undertaken to produce them. Reaume claims that such goods are *necessarily collective* in the sense of goods such that an agent that contributes to their production cannot be excluded from their enjoyment, either by another agent or by the agent himself. Examples are political participation and cultural participation. We might refer to such goods as *participatory* common goods.

Arguably, social institutions are an important means by which common goods that ought to be produced are in fact produced. Indeed, on some accounts of social institutions, the production (or maintenance or renewal) of common goods is the raison d'être of social institutions.

On Seumas Miller's account, the collective ends pursued by organized joint activity are common goods by virtue of their possession of the following three properties: (1) they are produced, maintained, or renewed by means of the *joint activity* of members of organizations such as schools, hospitals, welfare organizations, agribusinesses, electricity providers, and police services—that is, by institutional role occupants; (2) they are *available to the whole community:* for example, clean drinking water, a clean environment, basic foodstuffs, electricity, banking services, education, health, safety and security; and (3) they *ought* to be produced (or maintained or renewed) and made available to the whole community since they are desirable (as opposed to merely desired) and such that the members of the community have an *(institutional) joint moral right* to them.

Seumas Miller

See also Collective Moral Responsibility; Institutions as Moral Persons; Libertarianism, Political; Public Goods

Further Readings

Aquinas, T. (1965). *Summa theologiae* (T. Gilby et al., Trans.). New York, NY: McGraw-Hill.

Aristotle. (1981). *The politics* (T. A. Sinclair, Trans.). London, England: Penguin.

Green, T. H. (1883). *Prolegomena to ethics.* Oxford, England: Clarendon Press.

Head, J. G. (1974). *Public goods and public welfare.* Durham, NC: Duke University Press.

Hobbes, T. (2010). *The Leviathan* (Rev. ed.; A.P. Martinich & B. Battiste, Eds.). Peterborough, ON: Broadview Press. (Original work published 1651)

Miller, S. (2010). *The moral foundations of social institutions.* New York, NY: Cambridge University Press.

Olson, M. (1965). *The logic of collective action.* Cambridge, MA: Harvard University Press.

Raz, J. (1984). Rights-based moralities. In J. Waldron (Ed.), *Theories of rights* (pp. 182–200). Oxford, England: Oxford University Press.

Reaume, D. (1988). Individuals, groups and rights to public goods. *University of Toronto Law Journal, 38*. Retrieved from http://ssrn.com/abstract=1895704

Rousseau, J.-J. (1953). *The social contract* (M. Cranston, Trans.). London, England: Penguin. (Original work published 1762)

Taylor, M. (1987). *The possibility of cooperation*. New York, NY: Cambridge University Press.

Common Knowledge

While you are working in the library, the loudspeaker announces that the library will close in 15 minutes, inviting the public to move toward the exit. The public announcement makes you and everyone else in the library aware that the library is about to close for the day. You realize that all users in the library have heard the announcement and know that the library is about to close. The state of group knowledge in which all members of a group know that something is the case is called *mutual knowledge*.

The subject matter of this entry, common knowledge, is a far richer epistemic state than mere mutual knowledge. After hearing the announcement, you realize not only that everybody in the library knows that the library is about to close but also that everybody is in a position to come to the same realization, since the announcement was publicly made. Thus, everybody knows that everybody knows that the library is about to close. But everybody can realize also that everybody else can realize that the public announcement has made everyone aware that the library is about to close. Thus, everybody knows that everybody knows that everybody knows that the library is about to close. The chain of "realizations," inferences, awareness, and knowledge can, in principle, go on indefinitely. Thus, everybody knows that the library is going to close, everybody knows that everybody knows that, everybody knows that everybody knows that everybody knows that, and so on ad infinitum. This state of group knowledge is called *common knowledge*.

In this entry, we will describe the pervasiveness of common knowledge, offer two characterizations of it, and illustrate applications relevant to the social sciences.

Common knowledge is ubiquitous. As we have seen above, public announcements generate common knowledge. But common knowledge is also implied by the fundamental psychological mechanism of joint attention, and it is assumed in our linguistic exchanges: Definite references presuppose common knowledge, as does meaning in Gricean accounts (after the British philosopher Paul Grice). The philosopher who introduced the notion in his 1969 book on social conventions, David Lewis, argues that common knowledge is necessary for social conventions to exist. In the context of group agency and political obligation, Margaret Gilbert claims that common knowledge is necessary for a joint obligation to arise in a group of agents.

However, common knowledge is also elusive. It entails an infinite amount of epistemic states, which, of course, cannot possibly occur in finite agents. It entails perfect sharing of public announcements or events, which cannot always be assumed: Did the person who was listening to her iPod in the library hear the library loudspeaker announcement? It entails that agents be symmetric reasoners: If an agent infers p in a given situation, everyone else in that situation will infer p. Moreover, it requires that it be common knowledge that agents are symmetric reasoners. (This raises the question "Where does *that* common knowledge come from?") David Lewis ingeniously proposes that in our daily interactions common knowledge represents an infinite number of *potential* knowledge states, which translate into only a few steps of *actual* ones: In practice we seldom go over the first two or three levels of mutual knowledge.

The infinitary character of common knowledge can be characterized by a finite definition exploiting the circularity inherent in a publicly shared situation. A classic definition of common knowledge goes like this:

p is common knowledge in a group G if

1. E obtains,
2. E entails that everyone in G knows that E, and
3. E entails that everyone in G knows that p.

Consider the library example again: E is the announcement on the loudspeaker, p the event that the library will close in 15 minutes. It is easy to see that by repeatedly applying (3) to (2) and appropriately using

modus ponens, it is possible to derive the whole common knowledge hierarchy.

The above characterization of common knowledge can be compressed in the so-called fix-point account, used mainly by logicians and computer scientists. While common knowledge can be a valuable logical notion to use, most logical systems do not allow for infinitely long formulas, hence computer scientists investigating multi-agent architectures and philosophers investigating epistemic logic by and large base common knowledge logics on the fix-point account. In a nutshell, Conditions 1 to 3 above are summarized by a unique logical formula that, in English, reads "Everyone in *G* knows *p* and everyone in *G* knows **σ**," where **σ** refers to the whole expression between scare quotes.

Applications of common knowledge are widespread, and it will only be possible to give here three brief and partial references. First, in economics, one of the earliest and seminal results involving common knowledge is Robert Aumann's agreement theorem. Aumann proves that if two agents have a common prior probability distribution and they have common knowledge of their posterior distributions, then no matter how disparate the information on which they updated their priors, their posteriors are equal. The result has been generalized to no-trade theorems, showing the impossibility of speculative trade between rational agents. Second, in game theory and its philosophical foundations, much has been discussed about the role that common knowledge plays in the selection of action in strategic interactions between rational players: Cristina Bicchieri has developed this line of thought. Third, in the philosophy of social sciences, there is an ongoing debate on the role played by common knowledge in game-theoretic accounts of social convention. While authors influenced by the evolutionary approach stress that common knowledge is not necessary for conventions to arise, others argue that common knowledge is crucial to create expectations supporting coordinating action.

Giacomo Sillari

See also Collective Rationality; Conventions, Logic of; Cooperation/Coordination; Group Beliefs; Joint Attention and Social Cognition; Modal Logic and Intentional Agency; Mutual Beliefs; Social Conventions; Social Norms; Social Perception; Social Rules; Team Reasoning

Further Readings

Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics, 4*(6), 1236–1239.

Bicchieri, C. (1993). *Rationality and coordination.* Cambridge, England: Cambridge University Press.

Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. (1995). *Reasoning about knowledge.* Cambridge: MIT Press.

Gilbert, M. (2006). *A theory of political obligation.* Oxford, England: Oxford University Press.

Lewis, D. (1969). *Convention.* Cambridge, MA: Harvard University Press.

Common Sense (in the Social Sciences)

From the Latin *sensus communis,* common sense came to refer to a body of self-evident truths or first principles presupposed by the practical judgments of every sane individual. Deviating from this universal capacity, the philosophical understanding of common sense oscillated between ontology and epistemology, means and ends, judgment and knowledge, certainty and ambiguity, sensory experience and cognitive categories. Across philosophers as diverse as Aristotle, Giambattista Vico, Shaftesbury, Thomas Reid, Immanuel Kant, Henri Bergson, G. E. Moore, Ludwig Wittgenstein, Hans-Georg Gadamer, and Richard Rorty, common sense has been subject to competing, often terse, analyses and counteranalyses.

Sensus Communis

A humanist appeal is made to the sensus communis as the communal basis of reason, truth, and moral sense, in contrast to the convolutions of abstract theoretical knowledge. The Third Earl of Shaftesbury pointed to the lively wit and ridicule of the sensus communis as evidence of human sociability and morality. As it passed through the Scottish common sense philosophy of Thomas Reid, the universality of tacit common understanding represented the supreme court of appeal, while the philosophical "way of ideas" led ineluctably to "ridiculous" and "absurd" forms of "metaphysical lunacy." Reid argued that philosophers who cast doubt on our knowledge of reality, like David Hume, routinely adhere to commonsense principles in daily life.

Realist assumptions about universal common sense mean that individual experience cannot be deceived or doubted. Intellectual authority derives from direct observation and the combined senses of individuals bound together in community. Morally, all sane human beings are believed to be congenitally predisposed to the "good sense" of universal rules of conduct. Socially, common sense anchors individuals intersubjectively to others in specific social worlds. Psychologically, it provides ontological security to stabilize a unified sense of self caught up in the flow of life. Common sense also offers protection against the narrowing horizon of intellectual expertise generated by scientific overspecialization.

Duality of Common Sense

Social science adopts contrasting approaches to common sense. On one side, common sense is understood as a precondition of social solidarity and social action. Social science is thought to be continuous with common sense, an umbilical cord that can never be cut but simply needs to be more systematically ordered by self-conscious reason. On the other side, common sense is reproached as a constraint on critical reason and autonomy. Social science qua science strives for a discontinuous "break" with the self-evident biases and distortions of common sense.

With its emphasis on the social conditioning of sound judgment as a universal disposition, common sense entered the social sciences by a number of circuitous routes. In Britain and much of southern Europe, common sense possessed a critical, moral, and civic quality, underpinning rationalist and positivist assumptions in social thought. In the United States, commonsense philosophy was refracted through the Scottish school and pragmatism. "Critical commonsensism" formed a bridge between self, society, and science, helping found an alternative social psychology to John Locke's cognitive individualism. Commonsense universality provided the grounds for modern democratic challenges to arbitrary authority, from Tom Paine's 1776 pamphlet "Common Sense" to Martin Luther King Jr.'s 1963 "I Have a Dream" speech. The most celebrated economic tract of the 20th century, John Maynard Keynes's *The General Theory of Employment, Interest and Money* is indebted to the indeterminate possibilities of ordinary-language philosophy developed by his Cambridge colleagues G. E. Moore, Frank Ramsey, and Ludwig Wittgenstein.

This critical egalitarian understanding of common sense found little resonance in German metaphysics. Commonsense judgment was consigned to the lower functions of mind, languishing beneath an elevated scholastic cognitive tradition. For Immanuel Kant, the sensus communis provided the indeterminate ground for the universal judgment of aesthetic taste rather than specific truths of scientific reason or morality. Even in Enlightenment France, common sense came to be identified with the childlike credulity, prejudice, and superstition of the vulgar masses, in contrast to the enlightened reason, tact, restraint, and taste of cultivated intellectuals.

Common sense was excised from the pure analytical conceptions constructed by classical sociology. Émile Durkheim formulated a scientific "break" with the "previsions" of common sense by establishing a distinctive methodology and conceptual framework, from "social facts" to describe the impersonal norms of common sense and the "conscience collective" of highly integrated, premodern social systems. Drawing on the German hermeneutic tradition of Wilhelm Dilthey and Heinrich Rickert, Max Weber made subjective observation of purposive social action broadly continuous with common sense through the "sympathetic magic" of *aktuelle,* or "direct understanding." However, this was always subordinate to Weber's project to found an "objective social science" through ideal models of typical causes and motives of instrumental social action.

From Karl Marx's premise that being-in-the-world determines consciousness, Marxism found great difficulty assimilating common sense depending on whether social structures or sensual-practical subjectivity was emphasized. For Georg Lukács, empirical common sense represented a barrier to objective knowledge and class consciousness because it spontaneously reified the social world of universal commodity production. Louis Althusser reduced common sense to insidious ideological representations, from which only by a rigorous effort of scientific labor could one break free—exemplified by the later work of Marx.

In contrast, for Antonio Gramsci common sense represents the "folklore" of philosophy. Since it emerges among historically formed groups in specific cultural conditions, common sense for Gramsci (1971) represents "a chaotic aggregate of disparate conceptions" (p. 422). It is "ambiguous, contradictory and multiform" (p. 422) rather than a flat realm

of ideological transmission or reification. Common sense spontaneously provides the contradictory but fertile ground through which "the good sense" of class consciousness is channeled by the "philosophy of praxis" (Marxism).

Phenomenological Sociology

Synthesizing Edmund Husserl, Henri Bergson, pragmatism, and hermeneutics, Alfred Schütz developed a phenomenological approach to the "natural attitude" of common sense in the "life-world." For phenomenological sociology, common sense performs an axiomatic function that makes intersubjectivity possible in the first place. Routine, prereflexive common sense puts the social world "at hand" without much further ado as real, meaningful, and coherent. Phenomenological sociology is more concerned with the necessary preconditions of social worlds than with the explicit meanings or rationalizations of agents themselves, an insight developed independently by the later Wittgenstein. George Herbert Mead's "social theory of mind" acted as a pragmatic-behaviorist bridge to phenomenological sociology. For Mead the common socio-sensory capacities of individuals are essential for the completion of self by the "generalized other" through cooperation and conflict.

Phenomenological concerns with commonsense preconditions gave rise to a wide and varied intellectual tendency in postwar sociology, ranging from ethnomethodology, Erving Goffman's model of social action, the social construction of reality, through to feminist standpoint theories of gender, race, and class differences. Jürgen Habermas recentered critical theory's concern with the systemic loss of consensual good sense in the bourgeois public sphere through the malign construction of "public opinion."

Synthesizing phenomenology and classical sociology, Pierre Bourdieu developed a critique of scholastic disdain for common sense by theorizing the logic of *sens pratique* ("practical sense") as a concrete "feel for the game" of the embodied dispositions of habitus. This does not refer to a universal sense, as in the Kantian judgment of taste, or the contingent identities valorized by postmodernism, but to a common, prereflexive sense of unequal positions structured across social space.

Alex Law

See also Ethnomethodology; Habitus; Marxism and Social/Historical Explanation; Pragmatism; Scottish Enlightenment: Influence on the Social Sciences

Further Readings

Billig, M. (2008). *The hidden roots of critical psychology: Understanding the impact of Locke, Shaftesbury and Reid.* Thousand Oaks, CA: Sage.

Bourdieu, P. (1990). *The logic of practice.* Oxford, England: Polity Press.

Coates, J. (1996). *The claims of common sense: Moore, Wittgenstein, Keynes and the social sciences.* Cambridge, England: Cambridge University Press.

Gramsci, A. (1971). *Selections from the prison notebooks of Antonio Gramsci* (Q. Hoare & G. Nowell-Smith, Eds. & Trans.). New York, NY: International.

Reid, T. (1872). *The works of Thomas Reid* (Vols. 1 & 2). Edinburgh, England: Maclachlan & Stewart.

Schutz, A. (1962). *Collected papers 1: The problem of social reality* (M. Natanson, Ed.). The Hague, Netherlands: Martinus Nijhoff.

Communication Studies

Communication studies is a multifaceted area of academic inquiry in which scholars explore how humans share symbols to create meaning. The subject includes investigations into topics such as understanding, interaction, transferring meaning, and reducing uncertainty in various contexts. Communication studies is an interdisciplinary academic area, and theoretically applied topics of investigation frequently overlap with scholarship conducted in other academic disciplines including psychology, sociology, history, education, health care, political science, and business. While researchers in other academic disciplines consider communication to be a secondary process within their content area, communication scholars consider communication to be the primary content and other fields as secondary. Communication has a long history dating back to antiquity and includes a range of topics that may be examined from both humanities and social science perspectives.

History

Early Greek and Roman philosophers including Plato, Aristotle, Cicero, Quintilian, and others emphasized the importance of communication

in everyday life. They were interested in the use of persuasive arguments to influence others in the public sphere (*rhetoric*) and examined both in the practice of communication as well as various parts of the communication process. In fact, rhetoric and the logic of persuasion became a *topos* of systematic philosophical analysis in the hands of Aristotle but also a bone of major philosophical contention between the sophists, on the one hand, and Socrates and Plato, on the other—with important repercussions for ethics and the theory of knowledge.

One of the earliest models of communication codified the five canons of rhetoric. While some scholars emphasized the civic practice of combining these components to deliver a persuasive argument, other scholars engaged in the formal theoretical investigation of a particular segment of communication. Thus, each canon has a practical and theoretical aspect. Invention (Latin *inventio*) relates to identifying the content of an argument that one will make to other people and is informed by theoretical investigations into the nature of knowledge and knowing. Style (Latin *loqui*) refers to the language choices one selects to best convey ideas and is enriched by theories relating to the nature of language. Arrangement (Latin *dispositio*) considers how ideas are arranged or organized within one's speech and is enhanced by theories about the processes by which information could be ordered and related. Memory (Latin *memoria*) relates to one's ability to remember and recall information and is guided by theoretical content related to processes of information storage and retrieval. Delivery (Latin *pronuntiatio*) focuses on the recitation or presentation of the content and is enhanced by theories about message transmission.

Following the classical period, the medieval and Renaissance periods saw a decline in scholars' interest in formal communication theory. Communication scholars emphasized issues of practical concern in prescriptive and nontheoretical works, such as manuals on letter writing and preaching. The theoretical study of communication reemerged with the secularization of thought in the 1600s and included four major areas of inquiry. Classical scholarship on communication reintroduced and elaborated on the earliest works in rhetoric, reversing the prescriptive trend. Epistemological-psychological approaches raised questions about how humans know, believe, and act in different situations, emphasizing the mental processes involved in communication.

Belletristics (French *belle lettres,* "fine writing") opened up theoretical interest in the study of literature, poetry, and drama as communication, prompting the investigation of the relationship between language arts and broader society. Elocutionist scholarship emphasized the practical aspects of delivery, stimulating theory about the verbal and nonverbal behaviors one can use to enhance a presentation.

Contemporary Issues

The 20th century saw the rise of communication technologies, including radio, television, the telephone, satellites, and computer networking. This further broadened communication inquiry to include the use of mass communication channels to share information and seek to improve society. Along with scholars, members of society in general became interested in myriad communication topics, including communication as a psychological and social process, persuasion and influence, and decision making in groups. In keeping with societal changes, postmodern scholarship in communication studies is highly eclectic in both topics of investigation and the forms of scholarship used to investigate those topics.

Communication studies has been approached by scholars both as an area in the humanities and as a social science. The epistemological choice a scholar makes determines the research approach taken. Within the humanities, scholars consider the philosophical and rhetorical grounds of human communication, including topics such as meaning and truth, interpretation, representation, and speech acts. Humanities scholars may approach communication from various philosophical traditions (semiotics, hermeneutics, phenomenology, critical theory, etc.) or rhetorical traditions (various types of rhetorical criticism, including dramaturgical, genre, feminist, etc.). Approaching it as a social science, researchers employ both quantitative and qualitative methods to investigate communication topics such as attitude change, relationship development and deterioration, decision-making processes, and media effects. Some social science scholars choose to triangulate research methods and data analysis techniques—using quantitative or qualitative data-gathering techniques—to produce data that may be analyzed statistically or interpretively.

Communication studies is a term that includes a variety of topics that may be considered in any

number of ways. Domains of communication inquiry include intrapersonal, interpersonal, group, organizational, and public communication. Furthermore, communication in each domain may occur face-to-face or through a mediated channel in either private or public settings. Scholars examine how communication is symbolically structured and functions within each area. They also attend to factors such as a person or group's gender, ethnicity, culture, able-bodiedness, sexual orientation, age, ethics and religious value systems, and so forth. The forms of inquiry in communication studies are very broad and include the depth and breadth of subjects related to the human exchange of symbols to create shared meaning.

People have been interested in human communication since antiquity. The dynamics of communication include a variety of domains and range of topics that may be studied from a humanities/philosophical perspective or using social science methods.

Pat Arneson

See also Information Society; Language and Society; Semantics and Pragmatics; Speech Acts; Symbolic Interactionism; Symbolism

Further Readings

Arneson, P. (Ed.). (2007). *Perspectives on philosophy of communication.* West Lafayette, IN: Purdue University Press.

Cohen, H. (1994). *The history of speech communication: The emergence of a discipline, 1914–1945.* Annandale, VA: Speech Communication Association.

Littlejohn, S. W., & Foss, K. A. (2011). *Theories of human communication* (10th ed.). Long Grove, IL: Waveland Press.

Communicative Action Theory

Jürgen Habermas's theory of *communicative action* is central to his version of critical theory and a powerful alternative to theories of rational action and rational choice, which have been influential in the social sciences. It is also exceptionally relevant to this encyclopedia because it is located on the borders of philosophy and social theory.

In the 1970s, Habermas (1929–) had been attracted by the idea of using the theory of language developed in analytic philosophy and speech act theory as a basis for the social sciences. By 1981, when he published his *Theory of Communicative Action,* he had abandoned this aim, and he presents his theory as a social theory rather than as a metatheory. In its focus on a particular conception of rationality, however, it is, like rational action theory, taking up central philosophical themes. As Habermas stresses in the first chapter of *Theory of Communicative Action,* a concern with rationality is central both to philosophy and to the social sciences. In social science, and in a more informal way in everyday life, we make sense of other people's speech and behavior in terms of models of rationality ("I see why you said that"), we test these understandings against rational criteria ("But you completely misunderstood the situation"), and we make judgments about how far a person or a larger collectivity really has been rational.

Rather as the term *discourse* has been stretched far beyond a linguistic concern with speech and writing to include more complex systems of belief and practice (as in Michel Foucault's analyses of medical and psychiatric discourse), Habermas understands by communicative action a form of action oriented toward agreement, whether about states of affairs in the external world, normative claims about what should be the case, or proposals for action. Anyone engaging in communicative action is implicitly claiming that what he or she says is intelligible, true, sincere, and justified. Even a request, if it is more than just an instruction (like the pilot's order to the cabin crew to switch the "doors to automatic"), opens itself up to a possible process of justification. In Habermas's example in Volume 1 of *Theory of Communicative Action,* if he asks a seminar participant to fetch him a glass of water, he is making the implicit claims that there is a reasonably convenient source of water, that he really wants it and is not just attempting to belittle the student, and that the request is a reasonable and appropriate one that he is entitled to make; any or all of these claims may be contested by the addressee of the request.

In Volume 2, Habermas uses a similar example to make the point that communicative action is not just communication or a process of reaching understanding: It is also an ongoing element of sociation, engaging individuals' definitions and understandings of situations and their own identities. This time, the example is of an older building worker sending a

younger recent recruit to fetch beer to go with the midmorning snack. Again, the exchange makes sense against the background of a whole set of shared assumptions that may be problematized if the younger worker does not understand the custom or the expectations associated with his or her place in the team hierarchy.

Habermas makes the same sort of claim for the primordial character of communicative action as other theorists would make for self-interested or, in Habermas's terms, strategic action. Both, he insists in a reply to critics in 1986, have a teleological component. What distinguishes them is not so much the actor's orientation as such but what Habermas calls "the combination of actor attitudes (orientation to success vs. orientation to reaching understanding) and forms of coordinating different action plans (influence vs. consensus)." It is all right, in other words, for me to argue vigorously with you for my point of view about what is true or what is to be done about some matter of concern, but not for me to try to persuade you by rhetoric or manipulation, for example, by falsely suggesting that "there is no alternative."

Bargaining, manipulation, and so on, he argues, are all parasitic on communicative action, as are the sociological models of normatively oriented action (i.e., action shaped by a shared value system), derived from the functionalist sociologist Talcott Parsons, or the "dramaturgical" model of action developed by Erving Goffman, in which the idea of role-*playing* is taken seriously as a model of the "presentation of self" in social life. He was not, he insisted, saying that people choose or want to act communicatively but that they *have to*.

Armed with this model of communicative action, Habermas presented an outline history of Western modernity, following Marx and Max Weber, in which social arrangements came to be questioned in light of reason (the "Enlightenment"), only to be removed from the scope of rational discussion and possible agreement by the growth of capitalist markets and bureaucratic administration. There can be no argument if you have reached your credit limit or if your case is not eligible according to the rule book. This process of rationalization or reification, which for a time Habermas referred to as the "colonization of the life-world" by formal systems, might not have been inevitable: There might have been, or might still be, more communicative or cooperative alternatives.

Habermas went on to develop a *discourse theory of morality,* also less precisely called "discourse ethics," according to which norms are valid if all those affected could in principle come to agree on them. This is a theoretical justification as he presents it, but he also wants to claim that the more such discussion actually takes place and the closer it approximates to the ideal vanishing point, which for a time he called the "ideal speech situation," the better. In a society like ours, where there is no longer (if there ever was) a single shared view of the world and of how we should behave in it, the only way we can achieve a shared understanding is by bringing our interpretations into line with one another in a process that is "risky" but also promises greater rationality. There are echoes here of Max Weber's image, in which in the modern world we are faced by the rival demands of warring gods and demons; there are also philosophical echoes of Hegel's critique of Kant's abstract account of morality and his own—more sociological emphasis on the way it is necessarily embedded in the ethical life (*Sittlichkeit*) of a real human community.

Following a long engagement with legal theory, Habermas extended his model into a discourse theory of law and democracy, in which law gives teeth to morality but also demands democratic justification. Here, Habermas converged with theories of deliberative democracy as it was developed initially with reference to states and then, beyond state borders, in models of cosmopolitan democracy.

Critics of Habermas's model have tended to see it as excessively rationalistic, though it is worth noting that his friend Karl-Otto Apel's more Kantian version of discourse theory is more directly vulnerable to this charge. Approaches such as those of Michel Foucault or Pierre Bourdieu, which put more emphasis on power and strategic action, or analyses of political discourse, which give more attention to rhetorical and "agonistic" aspects, may seem to be more realistic. Analysts of discursive democracy, often focusing on deliberations in the European Union, have argued for the importance of acceptable compromises setting limits to, while still incorporating, the pursuit of self-interest. Habermas himself has no objection to majority voting, so long as a prior agreement has been reached that this is an acceptable way to decide an issue, for example, where pressure of time rules out further search for a consensus. Academic committees, for example, may

see voting as a last resort, whereas political ones may be happier to take a formal vote, even where there is unanimous agreement.

The distinction between communicative and strategic action may itself not be as clear-cut as Habermas assumes. The aesthetic dimension of social life also receives little attention in Habermas's model. Albrecht Wellmer, in particular, who had contributed substantially to the communicative action model, has come to stress this aspect, generalizing it to the argument that the "negative freedoms" of a liberal democracy should include the right "not to be completely rational." More concretely, some critics have detected a latent authoritarianism in the stress on the pursuit of agreement and the requirement of rational justification.

The early critiques along these lines by French intellectuals may have been misdirected, but the demands by religious and/or ethnic minorities for their "difference" to be respected without the need for formal justification pose a more substantial challenge to someone like Habermas, who has always been deeply respectful of different ways of life, even in relation to unpromising contexts such as the South of the United States before the Civil War. The title essay of *The Inclusion of the Other* (see Further Readings, below) is a partial response to these issues, as is his later engagement with questions of religious belief. Seyla Benhabib has taken up these themes in a number of important recent works.

William Outhwaite

See also Action, Philosophical Theory of; Discourse Analysis; Enlightenment, Critique of; Frankfurt School and Critical Social Theory; Life-World; Speech Acts; Transcendental Pragmatics

Further Readings

Benhabib, S. (2004). *The rights of others.* Cambridge, England: Cambridge University Press.

Habermas, J. (1981). *Theorie des Kommunikativen Handelns* (2 vols.). Frankfurt, Germany: Suhrkamp. (Translated by T. McCarthy as *The theory of communicative action,* Vol. 1, 1984, London, England: Heinemann; Vol. 2, 1987, Cambridge, England: Polity)

Habermas, J. (1991). *Vergangenheit als Zukunft* (2nd ed.; M. Heller, Ed.). Zurich, Germany: Pendo. (Expanded ed., 1993, Munich, Germany: Piper; Translated by M. Pensky as *The past as future,* 1994, Lincoln: University of Nebraska Press)

Habermas, J. (1992). *Faktizität und Geltung.* Frankfurt, Germany: Suhrkamp. (Translated by W. Rehg as *Between facts and norms,* 1996, Cambridge, England: Polity)

Habermas, J. (1996). *Die Einbeziehung des Anderen.* Frankfurt, Germany: Suhrkamp. (Translated as *The inclusion of the other,* C. Cronin & P. de Greiff, Eds., 1998, Cambridge, England: Polity)

Honneth, A., & Joas, H. (Eds.). (1991). *Communicative action* (J. Gaines & D. L. Jones, Eds. & Trans.). Cambridge, England: Polity. (Original work published 1986)

Outhwaite, W. (2009). *Habermas: A critical introduction* (2nd ed.). Palo Alto, CA: Stanford University Press.

Susen, S. (2007). *The foundations of the social: Between critical theory and reflexive sociology.* Oxford, England: Bardwell.

Wellmer, A. (1998). *Endgames.* Cambridge: MIT Press.

Complex Networks Theory and Social Phenomena

This entry introduces the emerging interdisciplinary field of complex networks, which incorporates insights from physics and mechanical engineering as well as computer science into an account of social networks and social interconnecting in general. Complex networks theory attempts to explain in mathematical and topological terms not only social networks but also social institutions (e.g., markets) or social practices (traffic) on the analogy of properties found in the natural world.

Background and Definitions

Complex networks theory is the study of physical systems and social phenomena represented as networks, with nodes representing the actors (entities) of the system and links encoding the interactions between pairs of nodes. With roots in mathematics, physics, computer science, and biology, complex networks theory marks a departure from the perception, previously widely held within those fields, that either random networks—links occurring between pairs of nodes with equal probability—or regular lattices were appropriate mathematical approximations of networks occurring in the natural world. A *complex system* can be informally defined as a system where components acting and interacting

according to simple rules results in nontrivial global behavior. The aspiration of complex networks theory is to understand and describe how the simple, local interactions of individual nodes and actors result in empirically observed intricate network structures and dynamics on multiple scales.

Structure

Complex networks theory grew out of two central studies, each one pointing out fundamental structural properties of real-world networks.

In 1998, Duncan Watts and Steven Strogatz showed that two interesting structural properties were both present in a number of real-world networks. First, all networks were so-called small worlds, where almost every node can be reached from every other node in a small number of jumps. This finding drew on earlier work by the American social psychologist Stanley Milgram and is reflected in the commonly known idea of *six degrees of separation*. Second, they showed that if we consider the neighbors of a typical node, we will tend to find more connections between them than one would expect at random; this property is known as *clustering*. Social networks, in particular, are known to be highly clustered: Two of your friends are more likely to be friends with one another than would be the case with two randomly chosen nodes.

Soon thereafter, studying the distribution of a number of connections per node, the node's *degree,* Albert-László Barabási and Réka Albert found that the degree distributions observed in many empirical networks tend to follow *power laws*. As a statistical distribution, the power law is radically different from the *Poisson distribution* that describes the connectivity of random networks; the power law degree distribution is characterized by a few highly connected nodes, sometimes called *hubs,* and a vast majority of nodes with low degree. These broad degree distributions have profound consequences for network properties, related to—for example—robustness and stability, epidemic threshold, and search.

Since these initial results, researchers have discovered structure on almost any scale between single nodes and the total network size. Starting from the properties of pairs of nodes, it soon became clear that in some networks, nodes of high degree systematically connect to nodes of low degree—this *disassortative* behavior increases overall robustness and is often seen in biological networks. Social networks tend to be *assortative*—nodes of high degree tend to link other nodes of high degree, while low-degree nodes tend to connect to other low-degree nodes. If we consider *motifs* of three to five nodes and the links between them, it turns out that some motifs occur much more frequently than one would expect at random. Other motifs never occur, although one would expect to see them often had nodes been linked to each other by chance. The specific motifs that are upregulated, as well as those that are suppressed, provide information about the function of the individual networks. On the meso scale—between motifs and the entire network—networks are typically organized into communities. *Communities* are sets of densely connected nodes—in social networks, they correspond to groups of friends, families, coworkers, and so on. Finally, many networks are organized according to a hierarchy that describes how the various structural elements are combined: how nodes are linked to form motifs, how motifs are combined to form communities, and how communities are joined to form the network itself.

Social Phenomena

Complex networks theory does not exist in a vacuum—networks have been studied in many other contexts. Starting with the work of Jacob L. Moreno in the 1920s, an entire branch of sociology, *social network analysis,* has been devoted to understanding social networks. In that larger context, the single component that truly distinguishes complex networks theory is *scale:* Fueled by analytical and computational methods from physics and computer science, complex networks theory has specifically added to our understanding of social phenomena, when it comes to understanding social systems at the largest scales.

Seen from this perspective, Watts's and Strogatz's strongest contribution consists not in revealing the presence of clustering in many networks and the small-world effect—both were previously known phenomena—but in showing explicitly that these effects are expressed in a global sense and in proposing quantitative explanations for why this is the case. Knowing that it is a "small world" is one thing, but understanding the overall network properties that are responsible for short path lengths and network navigability requires insight into the global structure of the system.

The power law degree distributions discovered by Barabási and Albert are naturally statistical descriptions of entire systems—which is the reason why their discovery is without a precursor in social network analysis. The consequences of the power law degree distributions are particularly important when expressed on the societal scale. For example, networks with power law degree distributions often occur in nature because this configuration is more robust to random node deletions. The effects of power laws also extend to dynamical processes; for example, the spread of an infectious disease displays very differently on a power law network than is expected by the standard assumptions of epidemiology. Many other network properties (e.g., communities) have impact processes occurring on the network.

While complex network theory remains a discipline primarily driven by empirical discoveries, recent developments in the theory focus primarily on two tasks: One central challenge is to strengthen the understanding of networks with many different types of nodes as well as multiple link types, potentially coexisting between any pair of nodes. The second central problem is to develop a unified framework for describing the dynamical aspects of complex networks—and how the dynamical aspects interact with the questions of *multiplexity* (the overlap of roles, exchanges, and affiliations in social relationships).

Sune Lehmann

See also Complexity; Complexity and the Social Sciences; Emergence; Social Networks; Systems Theory

Further Readings

Barabási, A.-L. (2002). *Linked: The new science of networks.* Cambridge, MA: Perseus.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512.

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world.* New York, NY: Cambridge University Press.

Jackson, M. O. (2008). *Social and economic networks.* Princeton, NJ: Princeton University Press.

Milgram, S. (1967). The small world problem. *Psychology Today, 1*(1), 60–67.

Newman, M. E. J. (2010). *Networks: An introduction.* Oxford, England: Oxford University Press.

Newman, M. E. J., Barabási, A.-L., & Watts, D. J. (Eds.). (2006). *The structure and dynamics of networks.* Princeton, NJ: Princeton University Press.

Watts, D. J. (1999). *Small worlds: The dynamics of networks between order and randomness.* Princeton, NJ: Princeton University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature, 393*(6684), 440–442.

Complexity

The term *complexity* derives etymologically from the Latin *plexus,* which means "interwoven." Intuitively, this implies that something complex is composed by elements that are difficult to separate. This difficulty arises from the relevant *interactions* that take place between components. This lack of separability is at odds with the classical scientific method—which has been used since the times of Galileo, Newton, Descartes, and Laplace—and has also influenced the fields of philosophy and engineering.

In recent decades, the scientific study of complexity and complex systems has initiated a paradigm shift in science and philosophy, proposing novel methods that take into account relevant interactions. At the same time, complexity is relevant to the social phenomena studied by social science. A number of issues in studying complexity, such as the question of irreducibility or systems theory, are directly relevant to social science.

This entry reviews the basic aspects of the notion of complexity, both scientific and philosophical, explains its uses and definitions, discusses the prospect of a science of complexity, and delineates the possibility of a "philosophy of complexity."

The Limits of Reductionism

Classical science and engineering have successfully used a reductionist methodology, that is, separating and simplifying phenomena in order to predict their future. This approach has been applied in a variety of domains. Nevertheless, in recent decades, the limits of reductionism have become evident in phenomena where interactions are relevant. Since reductionism separates, it has to ignore interactions. If interactions are relevant, reductionism is not suitable for studying complex phenomena.

There are plenty of phenomena that are better described from a nonreductionist or "complex" perspective. For example, insect swarms, flocks of birds, schools of fish, herds of animals, and human crowds exhibit a behavior at the group level that cannot be determined or predicted from individual behaviors or rules. Each animal makes local decisions depending on the behavior of its neighbors, thus interacting with them. Without interactions—that is, with reductionism—the collective behavior cannot be described. Through interactions, the group behavior can be well understood. This also applies to cells, brains, markets, cities, ecosystems, and biospheres.

In complex systems, having the "laws" of a system, plus initial and boundary conditions, are not enough to make a priori predictions. Since interactions generate novel information that is not present in initial or boundary conditions, predictability is limited. This is also known as *computational irreducibility;* that is, there is no shortcut to determine the future state of a system other than actually computing it.

Since classical (nonquantum mechanical) science and (versions of) philosophy assume that the world is predictable in principle, and relevant interactions limit predictability, many people have argued that a paradigm shift is required, and several novel proposals have been put forward in recent years.

The Complexity of "Complexity"

There is a broad variety of definitions of complexity, depending on the context in which the term is used. For example, the complexity of a string of bits—a sequence of zeroes and ones—can be described in terms of how easy it is to produce or compress that string. In this view, a simple string (e.g., "010101010101") would be easily produced or compressed, as opposed to a more "random" one (e.g., "011010010000"). However, some people make a distinction between complexity and randomness, placing complexity as a balance between "ordered" and "chaotic" dynamics.

A well-accepted measure of complexity is *the amount of information required to describe a phenomenon at a given scale.* In this view, more complex phenomena will require more information to be described at a particular scale than simpler ones. It is important to note that the scale is relevant to determine the amount of information, since, for example, a gas requires much more information to be described at an atomic scale (with all the details of the positions and momentums of molecules) than at a human scale (where all the molecular details are averaged to produce temperature, pressure, volume, etc.).

Complexity has also been used to describe phenomena where properties at a higher scale cannot be reduced to properties at a lower scale—in other words, when the whole is more than the sum of its parts (as in the notion of "emergence"). For example, a piece of gold has color, conductivity, malleability, and other "emergent" properties that cannot be reduced to the properties of gold atoms. In other words, there is a potentiality of novel behaviors and properties: A system with coordinated interacting elements can perform more complex functions than the independent aggregation of the same elements. Emergent properties cannot be reduced to the components of a system, because they depend on interactions. Thus, an approach to studying complex systems requires the observation of phenomena at multiple scales, without ignoring interactions. Formalisms such as multi-agent systems and network theory have proven to be useful for this purpose.

Complexity Science?

The scientific study of complexity, under that label, started in the 1980s. Some people argue that it is a science in its infancy, since it has been only a few decades since its inception and has yet to reveal its full potential. However, some people argue that complexity will never be a science in itself, because of its pervasiveness. Since complexity can be described in every phenomenon, a science of complexity would be too broad to be useful. A third camp defends the position that complexity is already a science in its own right. This debate certainly depends on the notion of what a science is.

Moreover, one can argue that all three viewpoints are correct to a certain degree. A scientific study of complex phenomena exists; this is not in dispute. People also agree that this study is offering new insights in all disciplines and has great potential, already yielding some fruits. The pervasiveness of complexity is also agreed upon. A scientific approach where interactions are considered (i.e., as nonreductionist) has been propagated in all disciplines. Whether or not people call this approach "complex" is irrelevant. The ideas and concepts of

the scientific study of complex systems are being propagated. Perhaps there will never be a science of complexity itself, but complexity is pushing paradigm shifts in all the sciences.

Toward a Philosophy of Complexity

Science has greatly influenced philosophy. For example, Newtonian physics led to philosophical materialism and mechanism. Perhaps unknowingly, the reductionist worldview stemming from classical physics seeped into ontology, where people argued that the only real phenomena were those described by physics—the laws of matter and energy—while all the rest were only epiphenomena, reducible to Newtonian dynamics in the times of Laplace and to elementary particles in recent decades.

Complexity has shown that reductionism is limited, in the sense that emergent properties cannot be reduced. In other words, the properties at a given scale cannot always be described completely in terms of properties at a lower scale. This has led people to debate on the reality of phenomena at different scales. This is crucial for social scientific explanations. For example, interactions are not necessarily describable in terms of physics, but they can have a causal effect on the physical world. An example can be seen with money, the value of which rests simply on an agreement between people; it is not describable in terms of physics. Still, money has a causal effect on matter and energy.

Complexity has also shown the relevance of the observer in the description of phenomena, since depending on the scale at which a phenomenon is described, its complexity will change. This is in contrast to classical epistemology, which seeks in objectivism the path to find the "true" nature of phenomena.

It should also be noted that the novel information generated by interactions in complex systems limits their predictability. Without randomness, complexity implies a particular nondeterminism characterized by computational irreducibility. In other words, complex phenomena cannot be known a priori.

An interesting feature of the philosophy of complexity is that it is very close to certain Asian philosophies, which were not influenced by reductionism and do consider the relevance of interactions. It could be said that Western philosophies are rediscovering the ideas of Eastern philosophies through the scientific study of complexity.

Carlos Gershenson

See also Complexity and the Social Sciences; Emergence; Models in Social Science; Reductionism in the Social Sciences; Systems Theory

Further Readings

Bar-Yam, Y. (1997). *Dynamics of complex systems: Studies in nonlinearity.* Boulder, CO: Westview Press.

Cilliers, P. (1998). *Complexity and postmodernism: Understanding complex systems.* New York, NY: Routledge.

Gershenson, C. (Ed.). (2008). *Complexity: 5 questions.* Birkerød, Denmark: Automatic Press/VIP.

Mitchell, M. (2009). *Complexity: A guided tour.* Oxford, England: Oxford University Press.

Complexity and the Social Sciences

Phrased in an encapsulated form, the essential element of the *complexity turn* across science can be summed up in the common English expression "The whole is greater than the sum of its parts." That is to say, complexity is concerned with emergence, with what physical scientists refer to as "the failure of superposition." The essence of emergence is that entities may have properties that apply to them as a whole and are not explicable in terms of properties of the components that constitute the entities.

This entry reviews the origin and meanings of the notion of complexity and its principal uses in current social theory.

Complexity

It is important not to confuse the idea of complexity with methodological holism. Holism of this sort asserts that there are properties that belong to entities as a whole and are not explicable other than by reference to the entity as a whole. Complexity considers that an essential source of the properties of an entity are the interactions among its components. In the most developed version of complexity theory, such interactions are considered to operate among the components of the entity, among the components with the entity as a whole, and among the components of the entity, the entity as a whole, other entities, and the components of other entities.

Complex entities are generally understood as being *systems* that possess the *emergent* properties. Emergence is not a new idea. Its origins can be traced back to G. H. Hughes in the 1880s, and it has informed unconventional positions in philosophy and the philosophy of science since that time. Emergence is a radical challenge to the *reductionist* program in scientific explanation. Reductionism seeks to explain the properties of more "complex" entities/systems solely in terms of the properties of simpler entities. Ultimately, reductionism seeks to explain all of reality, including social reality, in terms of fundamental physical properties of matter and energy, although of course this is not a real program in any science. Complexity does not deny the validity of the reductionist program in relation to simple systems, but rather, it argues that it cannot cope with complex systems with emergent properties and asserts that much of reality, not only social reality but also biological reality and in particular the intersections of the "natural" and the "social," can only be understood if we pay attention to emergence and develop a strategy of investigation that recognizes its implications. So complexity theory is fundamentally not a testable theory describing an aspect of reality but rather an ontology asserting the nature of most of reality itself.

Complex Systems

The history of complexity in social science is itself complex. It stems (a) in part from the tradition of systems theory in sociology associated with Talcott Parsons, although it is radically different in form; (b) in part from the General Systems Theory proposed by Ludwig von Bertalanffy; (c) in part from cybernetics and in particular from the debates associated with the Macy seminars of the 1950s, described by Katherine Hayles; (d) in part from developments in the physical and biological sciences that began with the turning of attention to ideas of *chaos* and the use of computing technology to explore deterministic chaos using *nonlinear* modeling and simulation approaches, often associated with the work of the U.S. Santa Fe Institute; and (e) in part from a rather different European scientific approach to chemical systems associated with the work of the physicist Ilya Prigogine.

All these approaches, with the exception of the general tendency, but not the full content, of Parsons's account, are concerned with systems understood as being *open* and *far from equilibric*. Such systems are open in the sense that they receive information and energy from their environment and pass waste and other information back to their environment. They are far from equilibric, not in the sense of general chaotic systems, where very small changes in the values of control parameters can generate massive differences in trajectories over time, but rather in the sense of possessing the capacity for radical change while maintaining their integrity. They can undergo *phase shifts*—qualitative changes of kind. The idea of trajectory is crucial. Complex systems are dynamic. They develop and change through time. The idea of phase shift—of qualitative transformation—can also be expressed by saying that change in them is nonlinear. This represents a considerable challenge to those quantitative social sciences, and in particular econometrics, that have developed linear models in which changes in system state are proportionate to changes in system inputs. Likewise, the notion of a far from equilibric system is contrary to neoclassical economics' emphasis on equilibrium.

Social Complexity: The Current Debate

Across the social sciences, there are a variety of current and often conflicting approaches to complexity. The sociologist Niklas Luhmann draws on the very strong epistemological arguments of the philosophically minded biologist Humberto Maturana, asserting that complex social systems are closed rather than open. In contrast, Paul Cilliers, while explicitly engaging with postmodernist traditions and in particular the work of the deconstructionist philosopher Jacques Derrida, argues for a much more fluid and open approach that sees complex systems as intersecting, with boundaries that are both real and socially constructed. This has much in common with the arguments for a synthesis of complexity and critical realism, advocated by David L. Harvey and David Byrne. It also resonates with the approach of Manuel DeLanda, who draws on the work of the philosopher Gilles Deleuze and in particular the idea of assemblage.

It is impossible to separate philosophical argument about the role of complexity in the social sciences from methodological debate, a debate in considerable part driven by the development of simulation approaches. R. Keith Sawyer goes so far as to argue that simulation through agent-based modeling represents a third wave of complexity, but

this position is rejected by those who see this kind of approach as a kind of comfort-zone scientism, corresponding to what Edgar Morin calls "restricted complexity," in contrast to a general-complexity frame of reference, which represents a radical break with the program of science as it has been understood since Newton. Simulations are plainly an important tool in exploring complexity, but a clear distinction has to be drawn between simulations that are based on purely abstract propositions and those (far less common in the literature) in which rules are calibrated using real data. In a general-complexity frame, the latter can provide empirically grounded narratives, but these have no greater status than any other form of narrative as scientific description. Good narratives will generally be multisourced, and part of the complexity program is a breaking down of the distinction between quantitative and qualitative scientific description and explanation, with these modes being understood as complementary rather than contradictory.

David Byrne

See also Agent-Based Modeling and Simulation in the Social Sciences; Complexity; Emergence; Emergence and Social Collectivism; Equilibrium in Economics and Game Theory; Reductionism in the Social Sciences; Systems Theory

Further Readings

Byrne, D. S. (1998). *Complexity theory and the social sciences.* London, England: Routledge.

Cilliers, P. (1998). *Complexity and postmodernism.* London, England: Routledge.

DeLanda, M. (2006). *A new philosophy of society.* London, England: Continuum.

Hayles, K. (1999). *How we became post-human.* Chicago, IL: University of Chicago Press.

Luhmann, N. (1995). *Social systems.* Stanford, CA: Stanford University Press.

Maturana, U., & Varela, F. (1980). *Autopoiesis and cognition.* Dordrecht, Netherlands: D. Reidel.

Morin, E. (2006). *Restricted complexity, general complexity.* Retrieved from http://cogprints.org/5217/1/Morin.pdf

Nicolis, G., & Prigogine, I. (1989). *Exploring complexity: An introduction.* New York, NY: Freeman.

Reed, M., & Harvey, D. L. (1992). The new science and the old: Complexity and realism in the social sciences. *Journal for the Theory of Social Behaviour, 22,* 356–379.

Sawyer, R. K. (2005). *Social emergence.* Cambridge, England: Cambridge University Press.

von Bertalanffy, L. (1968). *General systems theory.* Harmondsworth, England: Penguin.

Concepts

Human beings are makers and users of representations. Prominent among these are concepts, or the representations we employ in higher thought. When we formulate plans; construct theories, explanations, and narratives; deliberate over a course of action; and reason more generally, we are engaging in processes that manipulate concepts. Indeed, having concepts is necessary even to form a belief, a desire, or an intention. To believe that the earth revolves around the sun requires possessing the concepts *earth* and *sun*; to intend to sign a mortgage contract requires the concept *mortgage*; and to desire a gin and tonic requires the concept *gin*. Without possession of those concepts, and the means of combining them, thinking these thoughts would be impossible. Thus, concepts have often been termed *the building blocks of thought*.

This entry first looks into concepts considered as mental items in the head, then moves on to discuss the role of concepts in philosophical methodology (e.g., conceptual analysis), and will conclude by explaining vagueness, open texture, and essential contestability exhibited by some concepts. The latter feature, in particular, is one considered crucial in social theory.

Concepts and Categorization

Among psychologists, concepts are thought of as the mental representations that enable individuals to categorize objects in one way or another. For example, having the concept *elephant* enables one to recognize elephants as such; to sort elephants from other creatures, such as rhinoceroses and woolly mammoths; and perhaps to label elephants verbally. Categorization may be thought of as making a mental judgment that this thing is an F or as thinking that all of these Fs are Gs, as when one categorizes elephants as animals.

Psychologists who study concepts have focused on the structure of these representations and

how they are learned and processed. The classical theory of concepts assumed that concepts are like definitions: They consist of sets of necessary and sufficient conditions for belonging to a certain category. To have the concept *elephant*, *electron*, *prime number*, or *mayor* is to grasp such a definition. This view derives in part from the idea that understanding what it is to be a certain kind of thing involves grasping the essential features of that thing, which is what proper definitions are supposed to summarize.

The classical view fell out of favor in the wake of Willard van Orman Quine's attack on the existence of analytic definitions, as well as many empirical studies suggesting that people do not represent most categories in terms of necessary and sufficient conditions. Instead, people appear to treat many categories as if they were organized around similarity to a prototype. Prototypes are representations of the statistically typical or "normal" member of a category: The typical raven is black, but there are also albino ravens; typical tables are four legged, but there are also three-legged designs; and so on. When we think of ravens and tables, we often think of them in terms of these prototypical features, not in terms of a definition. Whether we take something to be a member of a category depends on how similar it is to the prototypical category member, not on whether it possesses the category definition. Similarity-based views of concepts can explain many facts about people's judgments, including the greater speed and accuracy with which people categorize prototypical category members, the existence of borderline cases, and the fact that certain inductive inferences are facilitated for more prototypical categories.

Another perspective on concepts says that they are organized around theories. The theory-based view (sometimes called the "Theory theory"—not to be confused with the "Theory Theory" of folk psychology, pertaining to the area of the philosophy of mind) says that concepts either *are* theories of a certain domain or are *embedded* in them. Theoretical concepts like *quark* get their meaning from the theories they are a part of (e.g., the Standard Model of particle physics), and so also, on this view, do ordinary concepts like *elm* and *computer*. A central part of these ordinary theories is that they are committed to psychological essentialism. *Essentialism* means that people tend to believe that categories possess hidden or unperceived essences that determine the kind of thing they are and that normally cause or produce their superficially observable traits. In distinction from the classical view, however, it is not assumed that we know what these essences actually are. Essentialist beliefs seem to emerge early: Children will judge that something is really a cat even if it is painted and scented to appear just like a skunk and will assume that something is a skunk if it has skunk parents, no matter whether it looks and acts like a skunk. Concepts belonging to many domains seem to go beyond simple perceivable prototypical features, contrary to what many similarity-based views would predict.

As this brief sketch indicates, the psychological role of concepts is a complex one, and explaining how we use them to categorize will require appealing to many types of processes, including abstraction, judging similarity, and causal-explanatory reasoning.

Concepts, Publicity, and Possession Conditions

Concepts are widely believed to be public, meaning that it is possible for many thinkers to possess the very same concept. This requirement of publicity goes back at least to Gottlob Frege, who argued that thoughts must be the kinds of things that can be grasped by many individuals and by the same individual at different times. Publicity is important because in offering psychological explanations we appeal to the thoughts that people have: Someone's behavior is explained by his or her having certain beliefs and desires, for instance. These explanations have a degree of generality built into them, in that anyone who thought the same would behave the same. So psychological explanations, insofar as they are intended to be general, assume that concepts and thoughts are public. In light of this, theories of concepts need to explain their possession conditions in a way that satisfies publicity.

Many theories of possession conditions are *inferentialist*. Inferentialism is the claim that possessing a concept is a matter of being able to draw the right sorts of inferences. For example, someone who possesses the concept *squirrel* can infer from how something looks and acts that it is a squirrel and infer from the fact that it is a squirrel to the fact that it is a certain kind of animal, that it caches and eats nuts, and so on. To have the concept *plus*, one

needs to be poised to infer that 2 plus 2 equals 4, as well as indefinitely many other arithmetical propositions. An *inferentialist* holds that what it is to have a concept is just to be poised to use it in making such inferences.

Inferentialists must tell us something about which inferences a person needs to be able to make in order to possess a concept. Here, there are two possible positions: *holism* and *localism*. Holistic theories claim that possessing any concept necessarily requires being ready to make many inferences involving that concept, and thus, possessing any concept requires possessing vastly many others. To have the concept *apple*, and thus to be able to think that something is an apple, one needs to understand, among other things, that apples are a certain kind of edible fruit, that fruits contain seeds and have a certain characteristic reproductive and life cycle, that they have a range of salient properties and uses, and so on. On holistic views, even grasping such a seemingly simple concept requires having a vast web of interrelated concepts and beliefs.

However, since concepts are individuated by their place in this web, any change anywhere necessarily alters some or all of the concepts that one possesses. Holistic theories have difficulty satisfying publicity, since almost no individuals will share all of their network of beliefs, and hence, no two individuals will share exactly the same concepts. A task for holistic theories, then, is to explain what it means for concepts to be similar enough to share across differing belief systems.

Localistic theories claim that possessing a concept requires being poised to draw some inferences, but not anywhere near as many as holists claim. Hence, localism says that possessing any concept requires having some others but not a large number. The paradigm examples here are well-defined concepts: Having *bachelor* requires only having *unmarried male*, since that is all that a bachelor is; the concept *lamp* might be exhausted by saying that it is *designed to give off light*. Perhaps while *red* is not a definable concept, one could not possess it without also possessing *color*. These concepts are ones that can be had simply by being poised to draw a few crucial or central inferences.

Localistic theories impose more stability than do holistic ones, since not every change in a person's beliefs also changes all of his or her concepts. However, localistic theories are also committed to the existence of analytic or conceptually necessary connections: These are the inferential links that are part of the very meaning of a concept. Many philosophers are skeptical of the existence of any such connections. As noted above, Quine argued forcefully that there is no reliable criterion for when we have a genuine conceptually necessary connection, as opposed to merely a deeply held empirical belief. The price of adopting localism is accepting the existence of strong analytic conceptual connections.

Atomistic theories take a different approach in opposition to inferentialism. They claim that concept possession is not based on the inferences one draws with a concept but rather with what that concept picks out in the world. Concepts for atomists are fundamentally a kind of category detector. A detector is a device that goes off or indicates to a creature when something is present in the environment. Having the concept *cat* or *red* is just a matter of having a mental detector that reliably tracks, covaries with, or otherwise carries information about the presence of cats or red things in the environment. Because these detectors can reliably inform a creature about the world around it, this approach is sometimes termed "informational semantics," and so atomism may be thought of as an information-based approach rather than an inferentialist one. Informational views and inferentialist views differ on whether the fundamental role of concepts is to detect categories in the environment or to facilitate inferences concerning categories.

The ability to reliably detect a category does not presuppose the possession of any other concepts in particular, so atomists do not need to posit the existence of conceptually necessary connections. Moreover, a detector for cats or red things can exist even if its connection to other concepts and beliefs changes; so atomism does not face the stability problem that holism faces. However, it is unclear whether many concepts are best thought of as detectors. Concepts like *recession, art, truth, quark, good, economic equilibrium,* and *negative number* are ones that it is hard to imagine simply detecting conditions in the environment, since they pick out highly abstract and theoretical categories that are by their nature difficult or impossible to detect.

Informational theories may be correct for some concepts then but not all of them.

Innateness and Concepts

The normal adult has a rich repertoire of concepts for living things and material substances, actions and other occurrences, tools and artworks, kinship and other social groupings, scientific and technical categories, and the framework of ordinary life in general. Infants and young children, however, possess very few of these, and certainly none that depend on language, acculturation, and pedagogy. Two important questions in understanding conceptual development are (1) What is the initial conceptual endowment that children begin with? (2) What processes exist to enrich that initial endowment?

The first question, in particular, arises in debates about nativism. Nativism itself is a contested notion, and there is no general agreement on what it means to call a psychological characteristic innate. However, it can be useful to understand claims of innateness as meaning that a characteristic is acquired by a non-psychological process—in particular, that it is not a *learned* characteristic. If it is agreed that learning is at least one way to enrich our conceptual repertoire, the first question becomes "Which concepts do we learn, and which do we acquire without needing to learn them?"

A classic answer advanced by empiricists such as John Locke and David Hume is that our innate concepts are just our sensory concepts—those that pick out immediately perceivable categories such as particular colors, textures, shapes, tangible and audible qualities, and so on. These only require the right sort of experience to be "awakened" in us; all our other concepts are then complex logical constructions out of these sensory primitives. *snowball*, for example, might be a complex representation partly consisting of *white, cold, round*, and *hard*. The empiricist ideal is to extend this model to all our concepts, such that everything in the mind is ultimately analyzable in terms of concepts derived directly from sensory experience. On this view, there would be relatively few primitive concepts and many complex ones.

Rationalists such as René Descartes and Gottfried Leibniz argue that empiricists are wrong to suppose that most of our concepts can be analyzed in sensory terms. Experience with the world may be needed to give us these concepts, but they are not, properly speaking, learned from these experiences. Modern-day rationalists like Jerry Fodor have turned this into an argument for the impossibility of learning any concepts whatsoever. Fodor's argument runs as follows: Assume that to learn anything is to acquire evidence in favor of a certain proposition. To learn that grass is green, for instance, is to find evidence that confirms this generalization about the natural color of grass. Learning, then, requires the ability to entertain this proposition. To do this requires having certain concepts—the concepts *grass*, *green*, and so on. But if this is so, we must already possess these concepts so that we can frame the proposition to be confirmed. Similarly, with the case of learning a new concept, there must be some hypothesis involving the concept that we are trying to confirm. But going by the reasoning given above, any such hypothesis must be represented by the learner; which is to say that the learner must already have the concept that she is trying to learn in order to frame the hypothesis in the first place. So there cannot be any such thing as learning a concept, since any such act of learning presupposes the concept to be learned. This leads to a kind of radical concept nativism.

The contemporary debate between defenders of empiricism and rationalism has two major axes. First, there is the issue of how much of our conceptual repertoire is "copied" from sensory experiences. Neo-empiricists argue, by appeal to psychological and neuroscientific evidence, that all our concepts are ultimately grounded in the senses. If they are right, higher thought turns out to be a complicated form of perceptual simulation. Their opponents maintain that most, if not all, of our higher thought takes place in an amodal system that is independent of these lower-level perceptual systems. This remains a live question, in part since the notion of what should count as evidence that a psychological process is "perceptual" versus "amodal" is unclear and contested.

Second, there is the issue of whether we can learn any genuinely new concepts. Advocates of concept nativism think that the very idea of such learning is incoherent. Their opponents hold that the notion of learning as hypothesis confirmation is unnecessarily restrictive. They propose a variety of other models

on which we can acquire new concepts in a way that is appropriately adaptive and responsive to experience but that does not assume that we already have the concepts in question. These "bootstrapping" processes might treat concept learning as akin to learning a new skill or adding a new vocabulary item to a language. Neither of these obviously involves hypothesis confirmation in any problematic way. It remains open whether learning a concept is relevantly similar to these processes and whether they can avoid the radical nativist's argument.

Conceptual Analysis and Folk Concepts

Concepts have played a significant role in philosophical methodology. A central tenet of so-called analytic philosophy is that the proper task of the philosopher is to analyze important concepts such as *cause, action, person, knowledge, free will, truth, meaning, good, justice*, and so on. A common goal of philosophical analysis is the discovery of noncircular necessary and sufficient conditions for falling under a concept. Hence, the traditional analysis of *knowledge* claims that knowing that *p* involves having a justified true belief that *p*. Analyses of this kind play an important role in philosophical debate.

Analyses of concepts are often tested against thought experiments: real or fictional scenarios in which certain properties that are part of the analysis are present. In such scenarios, one asks whether the concept being analyzed properly applies in that situation or not. If not, then this shows that the analysis is somehow defective, since something may have the properties picked out by the analysis but not have the concept being analyzed. For instance, if *knowledge* simply is *justified true belief*, nothing could fall under the latter and not under the former. But as Edmund Gettier showed, seemingly decisively, there are cases of justified true belief that are not also cases of knowledge, and so the traditional analysis must be defective.

In recent years, experimental philosophers have challenged this method of doing philosophy. They have argued that the judgments that are elicited by these thought experiments are much less stable than philosophers have normally assumed. For instance, some of them show significant cross-cultural variation between Asian and Western populations, while others may vary across gender, degree of philosophical experience, and so on. Many conceptual analysts have assumed that their own intuitive judgments are representative of "our" concepts in general. But if these judgments are highly variable and unstable, this assumption is false. The "folk" may not think the way philosophers have assumed that they do or the way philosophers themselves do. If this is true, then concepts may not be as public as has often been thought; and if philosophers remain interested in analyzing concepts, they must decide which population's concepts are of interest and sample them more systematically. Seen in this light, there may be little reason to treat the concepts and intuitions of philosophers as being of any special interest. One response would be to continue to pursue the project of conceptual analysis but to broaden it to cover a more inclusive set of empirically gathered intuitive judgments. Alternatively, philosophers might adopt a conception of philosophy that does not see it as primarily being in the business of analyzing concepts by means of thought experiments.

Open Texture, Vagueness, and Contestedness

Many of our concepts are neither sharp nor precise. This imprecision comes in several forms. Friedrich Waismann argued that most terms in natural language are "open-textured," by which he meant that the rules or conditions that govern their application do not determinately cover every possible case. We might encounter things that have some of the characteristics associated with falling under a concept, such as *cat*, but others that are so irregular or eccentric that we are unsure whether to count them as category members or not. (Imagine something that looks and acts like an ordinary cat but that keeps growing to a gigantic size or appears and disappears seemingly at random or occasionally transforms into something doglike.) While for ordinary cases questions of category membership seem settled, there are indefinitely many cases where our concepts simply do not specify how we are to "go on." The notion of open texture was also famously invoked by the legal theorist H. L. A. Hart to describe the way in which laws necessarily fail to cover every possible future case that might fall under them (e.g., "No vehicles in the park" depends on what may count as a vehicle in the future—are Segways and hoverbikes included?).

Vagueness exists where there are borderline cases of category membership. Whether a person is tall

or bald is a vague matter, since while there are clear end points where membership is determinate, there are also many cases where it is not. A person with no hair is clearly bald, but many people are shy of a full head of hair, yet they are obviously not bald. We may want to say that there is no fact about whether they are bald or not or that the judgment that they are bald does not have a truth-value. Vagueness and open texture differ in subtle ways. Both involve indeterminacy of application and, hence, involve borderline cases. Vagueness in a concept can be eliminated by adopting some precise standard that fixes the boundaries of its application on a particular occasion of use, whereas open-textured concepts have indefinitely many possibilities for indeterminacy in their application, and these possibilities are unpredictable in advance. In other words, we usually know the dimensions or features that must be made more precise to resolve vagueness, but for open-textured concepts, we have no set list of qualities to make precise.

A different kind of demarcation problem is raised by what the political theorist W. B. Gallie referred to as "essentially contested concepts." Essentially contested concepts possess five features: (1) they are used to praise activities or achievements, (2) these achievements have internally complex features, (3) these features can be weighted in several possible ways to determine how the achievement is assessed, (4) the way these features can be modified or weighted is empirically sensitive and open-ended, and (5) the concept plays a dialogic role in which each user of the concept knows that his or her use will be contested and challenged by other concept users.

Gallie's examples of essentially contested concepts include *art, democracy, social justice*, and *christian life*. All of these denote a certain kind of complex, multidimensional activity that possesses the relevant open-ended empirical character, but one where the component dimensions may be assigned different weights by different people. Moreover, these concepts have both aggressive and defensive uses, with partisans of differing value schemes pitted against one another. Consider debates over whether Marcel Duchamp's readymades, Jackson Pollock's action paintings, or Damien Hirst's "The Physical Impossibility of Death in the Mind of Someone Living" (a tiger shark floating in a glass and steel tank full of formaldehyde) fall under *art*. In legal philosophy and the social sciences, the notion of essential contestedness has found wide application. Further widely discussed examples include concepts such as *rule of law, power, racism*, and *medicine*.

While essential contestedness is similar to open texture and vagueness, there are important differences. Like cases of open texture, essentially contested concepts do not have their empirical domain of application settled in advance. There is room for new exemplars to alter the extension of the category that they pick out. Thus, John Cage's *4'33"* may change what counts as a composition. On the other hand, open-textured concepts typically have a single, unitary, and widely agreed-on core of application. Essentially contested concepts do not, or at least need not. Partisans of two notions of *art* or *democracy* need not agree on the central exemplars of those concepts, since they may give the relevant evaluative features different weights.

In terms of the psychological theories of concepts discussed earlier, vagueness can be captured by many similarity-based theories of concepts, since these inherently allow for borderline cases. Essentially contested and open-textured concepts, on the other hand, have a complex internal structure reminiscent of the theory-based view. The essentially contested nature of many philosophical concepts may also explain the fact that traditional philosophical debates seem so peculiarly intractable.

Daniel A. Weiskopf

See also Analytic/Synthetic Distinction; Essentialism; Experimental Philosophy; Holism, in the Philosophy of Language; Inferentialism; Intentionality; Kinds: Natural Kinds Versus Human Kinds; Nonconceptual Content; Rule Following; Teleosemantics

Further Readings

Alexander, J., & Weinberg, J. (2007). Analytic epistemology and experimental philosophy. *Philosophy Compass, 2,* 56–80.

Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy, 10,* 615–678.

Carey, S. (2009). *The origin of concepts.* Oxford, England: Oxford University Press.

Cowie, F. (1999). *What's within? Nativism reconsidered.* Oxford, England: Oxford University Press.

Devitt, M. (1995). *Coming to our senses: A naturalistic program for semantic localism.* Cambridge, England: Cambridge University Press.

Fodor, J. (1981). The present status of the innateness controversy. In *Representations* (pp. 257–316). Cambridge: MIT Press.
Fodor, J. (1998). *Concepts: Where cognitive science went wrong*. New York, NY: Oxford University Press.
Fodor, J., & Lepore, E. (1992). *Holism: A shopper's guide*. Cambridge, MA: Blackwell.
Frege, G. (1956). The thought: A logical inquiry. *Mind, 65,* 289–311.
Gallie, W. B. (1956). Essentially contested concepts. *Proceedings of the Aristotelian Society, 56,* 167–198.
Gettier, E. (1963). Is justified true belief knowledge? *Analysis, 23,* 121–123.
Hart, H. L. A. (1997). *The concept of law* (2nd ed.). Oxford, England: Oxford University Press.
Keefe, R., & Smith, P. (Eds.). (1997). *Vagueness: A reader.* Cambridge: MIT Press.
Knobe, J., & Nichols, S. (Eds.). (2008). *Experimental philosophy.* Oxford, England: Oxford University Press.
Margolis, E., & Laurence, S. (Eds.). (1999). *Concepts: Core readings*. Cambridge: MIT Press.
Murphy, G. (2002). *The big book of concepts*. Cambridge: MIT Press.
Nagel, J. (in press). Intuitions and experiments: A defense of the case method in epistemology. *Philosophy and Phenomenological Research*.
Peacocke, C. (1992). *A study of concepts*. Cambridge: MIT Press.
Prinz, J. (2002). *Furnishing the mind*. Cambridge: MIT Press.
Quine, W. V. O. (1953). Two dogmas of empiricism. In *From a logical point of view* (pp. 20–46). Cambridge, MA: Harvard University Press.
Waismann, F. (1951). Verifiability. In A. Flew (Ed.), *Logic and language* (pp. 122–151). New York, NY: Doubleday.
Waldron, J. (2002). Is the rule of law an essentially contested concept (in Florida)? *Law and Philosophy, 21,* 137–164.

Consciousness

The term *consciousness* applies to several distinct phenomena, each pertaining to aspects of the mental functioning of people or other creatures. We describe people and other creatures as *conscious* when they are awake and sentient, as against being asleep, knocked out, or comatose. We also use the word *conscious* to describe a creature as being conscious or *aware of* something. And we describe thoughts, feelings, desires, and perceptions as being conscious in contrast with *unconscious* states of those sorts. All these phenomena have been the subject of extensive recent investigation in philosophy, experimental psychology, and neuroscience and the subject also of some heated debate in those fields. This entry discusses these types of consciousness and the major approaches to understanding them.

Awakeness and Awareness

An individual's being conscious, as against being asleep, anesthetized, and so forth, consists in that individual's being awake and receptive to sensory input and perhaps also having the ability to engage in voluntary, purposive movement. Investigation of this type of consciousness primarily concerns what aspect of *neural* functioning determines whether an individual is conscious or not. Understanding the neurological difference between conscious and unconscious conditions is important for determining the effects of various methods of anesthesia and for the clinical evaluation of patients in an apparent vegetative state, who have no discernible voluntary control of movement.

An individual when conscious has at least the ability to sense things and to have desires or intentions. Some argue that no creature is conscious in this way unless it actually senses or perceives, and perhaps such sensations or perceptions must be conscious states. But psychologically primitive creatures, such as snakes, may sometimes be awake and have the ability to sense things even when no actual sensing is taking place. It is likely that the conditions for a creature's being conscious cannot be settled except by appeal to reliable neurological correlates for such consciousness, established by appeal to uncontroversial cases.

Being conscious or aware of something often consists in an individual's sensing or perceiving that thing. If a person sees, hears, or smells an object, the person is conscious of that object. But even if one does not sense or perceive something, having a thought about that thing is also a way of being conscious of it, at least if one thinks about the thing as being present. It is less natural to describe somebody as conscious of something if the thought represents the object as distant in time or space.

There are cases in which a person perceives something *subliminally*—that is, without being aware of

perceiving it. Subliminal perception is perception that is below the threshold of conscious awareness. There is dispute about the correct way to describe such cases, but it is natural to think that in subliminal perception one is aware of the stimulus but not consciously aware of it. Such cases have been extensively investigated in experimental psychology, mainly by presenting individuals with stimuli that are degraded in some way—for example, they are very brief or faint—or stimuli that occur just before or after another stimulus that blocks conscious awareness of the target. The stimulus that does the blocking in such cases is said to be a *mask*.

Participants in these experiments report consciously seeing the mask but not the masked stimulus, though when the mask does not occur, participants do report consciously seeing that stimulus. But despite participants' denial that they see such masked stimuli, there is evidence that they do see it, since masked stimuli often have an effect, known as a *priming effect,* on subsequent psychological processing. Thus, participants may make a subsequent choice or have a faster reaction time in ways that reflect the character of the masked stimulus, despite claiming to be unaware of it.

Mental States and Consciousness

Despite evidence that perceptions occur in these subliminal cases, individuals report being unaware of any such perceptions and deny having them. So it is natural to describe those perceptions themselves as not being conscious. A conscious mental state, by contrast, is then a state an individual is aware of being in and can under ordinary circumstances report.

Subliminal perception is not the only case of mental states that individuals are unaware of. The best explanation for people's behavior is sometimes that they have acted on beliefs and intentions they deny having; experimental psychologists in these cases may posit unconscious beliefs and intentions. There is even evidence from social psychology that people sometimes do not have the beliefs and desires they claim explain their behavior. Rather, people confabulate having beliefs and desires that make sense of their behavior or fit with others' expectations or preconceptions.

There is also neuropsychological evidence that the neural events thought to correspond to decisions to do things occur before people are aware of making those decisions, at least one third of a second, and perhaps significantly more, in advance. This suggests that decisions occur and exert their causal influence in guiding behavior before one comes to be aware of the decision—that is, before the decision becomes conscious.

These findings show that mental states of all sorts occur without being conscious. So they encourage the view that a mental state's being conscious requires that one be aware of that state, and that if one is not so aware, the state is not conscious. This view of what it is for a thought, feeling, or perception to be conscious fits well, moreover, with our everyday understanding about mental states' being conscious. We sometimes take ourselves to see that another person thinks or wants something or feels a particular way, even though that person is unaware of doing so; those thoughts, desires, and feelings are unconscious.

Holding that a mental state is conscious *only if one is aware of it* also fits comfortably with psychoanalytic theory. The unconscious states that Freudian theory and its descendants posit are states that there is clinical reason to believe an individual is in despite the individual's being unaware of them. It is reasonable also to see the actions performed under post-hypnotic suggestion as due to volitions that the individual is again unaware of.

The idea that a mental state's being conscious requires that one be aware of that state is reflected in what psychologists call a *subjective measure* of consciousness, which holds that a mental state is conscious if it seems subjectively to one that one is in it. Some psychologists operate instead with an *objective measure* of consciousness, on which a state is held to be conscious only if there is evidence, independent of any subjective impression, that the state occurs. On this measure, an individual consciously perceives something only if that individual behaves in ways that reflect such perceiving. It is unclear, however, whether an objective measure can accommodate the occurrence of perceiving that is not conscious, since that would likely result in the same behavior as conscious perceiving. Another difficulty with the objective measure is that we can perceive things peripherally and also consciously, although such perceiving may have no measurable effect on behavior.

Qualitative Consciousness

It is reasonably natural to see people as having thoughts and volitions that are not conscious; actual and likely behavior points to their occurrence. But many profess difficulty in understanding how sensations, perceptions, and other states that exhibit *qualitative* character ("It feels like …") could occur without being conscious. That may be partly because an inference from behavior to qualitative states seems less compelling. But many theorists also regard qualitative states as having a special tie to consciousness; indeed, the term *consciousness* is sometimes used simply to refer to conscious *qualitative* states.

In this spirit, Ned Block has posited a type of consciousness distinctive of qualitative states, which Block calls "phenomenal consciousness." Phenomenal consciousness occurs when there is something it is like to be in a mental state, as with conscious perceptions, bodily sensations, and emotions. Block distinguishes this from what he calls "access consciousness," which consists in the representational content of a mental state's being available for the control of action, speech, and rational thought. When an individual cannot report being in a qualitative state, Block argues, it shows only that the state lacks access consciousness, not that it lacks phenomenal consciousness as well.

Some in philosophy have urged that special problems affect our understanding of conscious qualitative states. Even if we had accurate correlations of particular types of mental quality with types of neural state, we might be unable to explain why one type of neural state should occur in connection with a particular type of mental quality, or indeed any mental quality at all. Such an explanatory gap would presumably occur only for qualitative states that are conscious; if qualitative states can occur without being conscious, nothing would impede explaining such correlations.

It is also held by some in philosophy that there is no way to tell whether the mental quality that characterizes one person's seeing a red object is the same as the mental quality that occurs when another person sees such an object. Perhaps one person's mental quality when seeing red is the same as another person's on seeing a green object (this goes back to John Locke's hypothesis for the possibility of an undetectable "inverted spectrum of colors"). This concern stems from thinking of mental qualities as fixed solely by the way they present themselves to conscious awareness; otherwise, we could tell in some objective way whether such interpersonal quality inversion occurs. So this concern also pertains only to qualitative states as they occur consciously.

Perceiving prima facie involves mental states with mental quality. So evidence for subliminal perception puts pressure on the denial that qualitative states must always be conscious. Theorists who deny that qualitative states can occur without being conscious might urge that subliminal perceptual states never fail altogether to be conscious; they might maintain, for example, that subliminal states lack only access consciousness but not phenomenal consciousness. But it is arguable that subliminal perceptions also lack phenomenal consciousness, since there is nothing it is like for one to be in a state that one is wholly unaware of.

Instead of denying that subliminal perceptions fail to be conscious, a theorist who holds that qualitative states are always conscious can instead deny that subliminal states involve qualitative character. Perhaps they are simply neural states and not mental at all. But that is unlikely. Subliminal perceptions affect subsequent psychological processing in ways that appear to reflect differences in qualitative character, much as conscious perceptions do. It is natural to conclude that subliminal states have qualitative character and so qualitative states are not invariably conscious. Phenomenal consciousness is simply the special case of mental states being conscious where the states exhibit mental qualities.

The insistence that qualitative mental states are always conscious stems from the view that the very nature of mental qualities is determined solely by how they present themselves to conscious awareness. This insistence is captured vividly in Frank Jackson's famed thought experiment of somebody knowing everything about neural function and color perception but having never seen red and so not knowing what it is like to see red (though she knows the whole gamut of chemical theories about colors, neurological theories of perception, etc.). But knowing what it is like to see red is simply consciously having that experience. So the thought experiment gives no independent reason to hold that the nature of mental qualities is solely a matter of how they present themselves to conscious awareness. And subliminal perception shows that we know about

mental qualities not just by way of consciousness but also by the role they play in perceiving; mental qualities are the mental properties in virtue of which we sense and perceive things. Since we perceive both consciously and subliminally, there is no mystery about the occurrence of qualitative mental states that are not conscious.

Theories of Consciousness

Those impressed by evidence that mental states occur without being conscious tend to favor a theoretical approach that can explain the difference between mental states that are conscious and those that are not. One widely adopted approach builds on the idea that a mental state one is wholly unaware of is not a conscious state. A state is conscious, on this approach, only if one is aware of that state; theories that explain in this way what it is for a state to be conscious are known as *higher-order theories*. An apparent problem with such theories is that we seldom seem subjectively to be aware of our conscious mental states. Higher-order theorists usually argue that this is because the higher-order states by virtue of which we are aware of conscious states are seldom conscious states, so we are seldom aware of those higher-order states.

Another type of theory explains how conscious states differ from unconscious mental states by appeal to the effect that conscious states can have on psychological processing but that unconscious states cannot. On these *global-workspace theories,* as with Block's notion of access consciousness, a state's being conscious does not involve any awareness of the state; rather, a state is conscious if it is accessible to many significant types of psychological processing. And as with Block's notion of access consciousness, such theories face a difficulty similar to that which is faced by objective measures of consciousness, that peripheral perceptions can be conscious but have no significant effect on psychological processing. It may also be that repressed mental states, though not conscious, often do have a significant effect on psychological processing.

Theories like global-workspace theories, on which a state's being conscious does not involve any awareness of the states, are *first-order theories*. Other first-order theories maintain that a state's being conscious consists simply in its making one aware of something; on still others, a state is conscious if it involves attention; in no case does a state's being conscious require awareness of the state itself. Unlike global-workspace theories, it is unclear whether these types of first-order theory can accommodate mental states that are not conscious, as in subliminal perception. And it has been shown experimentally that states that are not conscious can involve attention and that attention is lacking in many conscious states, for example, conscious peripheral perceptions.

There are several versions of higher-order theory, which differ principally on what kind of higher-order awareness figures in states' being conscious. The theory prevalent in traditional philosophy is the *inner-sense theory,* on which a state is conscious if one is aware of it by sensing or perceiving it. Since none of the ordinary sense modalities can figure in such higher-order perceiving, a special inner sense is posited. Inner sense explains the apparent immediacy of our awareness of our conscious states, since perceiving and sensing always seem unmediated. But it is doubtful that any sense modality could make us aware of all the types of mental state that occur consciously, including perceptual states of different sense modalities. Nor is it clear what mental qualities could characterize such an all-purpose modality of inner sense.

On an alternative theory, we are aware of conscious states by having *thoughts* about those states. This avoids the problems facing inner sense, since thoughts can be about mental states of whatever kind and thoughts involve no mental qualities. This *higher-order-thought theory* explains the apparent immediacy of the way we are aware of conscious states, by hypothesizing that higher-order thoughts occur without any conscious mediation, for example, without relying on conscious inference or observation.

Another type of higher-order theory seeks to accommodate the considerations that motivate first-order theories by positing that the higher-order awareness by virtue of which mental states are conscious is *internal* to the conscious states themselves. Such theories must provide an independently motivated way of individuating mental states, on which such higher-order awareness is internal to the states we are aware of. It is unclear whether any method of individuation that had that result could square with the finding that many mental states at least occur measurably in advance of our becoming aware of them.

A theory of consciousness should address what *function* there is for a creature's mental states to be

conscious. It is natural to expect that there is some *utility* to a mental state's being conscious, since so many mental states are conscious. But we must distinguish the utility of a state's being conscious from the utility of states that happen to be conscious. When a conscious state has utility, that utility might be due not to the state's being conscious but to its other mental properties, such as what it represents or causes independently of its being conscious. So care must be taken in determining whether there is significant utility for mental states' being conscious.

David Rosenthal

See also Agency; Intentionality; Psychoanalysis, Philosophical Issues in; Self and Essential Indexicality; Self-Knowledge; Social Anti-Individualism and the Mental; Unconscious

Further Readings

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18,* 227–247.

Breitmeyer, B. G., & Öğmen, H. (2006). *Visual masking: Time slices through conscious and unconscious vision.* New York, NY: Oxford University Press.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory.* New York, NY: Oxford University Press.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workshop framework. *Cognition, 79,* 1–37.

Dienes, Z., & Seth, A. (2010). The conscious and the unconscious. In G. F. Koob, M. Le Moal, & R. F. Thompson (Eds.), *Encyclopedia of behavioral neuroscience* (Vol. 1, pp. 322–327). Oxford, England: Academic Press.

Dretske, F. (2000). *Perception, knowledge, and belief: Selected essays.* Cambridge, England: Cambridge University Press.

Freud, S. (1953–1974). The unconscious. In J. Strachey (Ed.), *The complete psychological works of Sigmund Freud* (Vol. 14, pp. 166–215). London, England: Hogarth Press. (Original work published 1915)

Gennaro, R. (Ed.). (2004). *Higher-order theories of consciousness.* Philadelphia, PA: John Benjamins.

Hassin, R. R., Uleman, J. S., & Bargh, J. A. (Eds.). (2005). *The new unconscious.* New York, NY: Oxford University Press.

Jackson, F. (1998). *Mind, method, and conditionals: Selected essays.* London, England: Routledge.

Levine, J. (2001). *Purple haze: The puzzle of consciousness.* New York, NY: Oxford University Press.

Libet, B. (2004). *Mind time: The temporal factor in consciousness.* Cambridge, MA: Harvard University Press.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Rosenthal, D. (2005). *Consciousness and mind.* Oxford, England: Clarendon Press.

Contemporary French Philosophy and the Social Sciences

Contemporary French philosophy has long been associated with the names of Jacques Derrida, Michel Foucault, Louis Althusser, Gilles Deleuze, Michel Serres, and Jean-François Lyotard (to which the name of Jacques Lacan, a psychiatrist and psychoanalyst exceptionally well-read in philosophy, has often been added), to name only the most renowned ones starting to flourish in the 1960s. There were close and complex interactions among these thinkers, often dating back to their education at the École Normale Supérieure (ENS), a most selective higher-education establishment near the Sorbonne and the French equivalent of an Oxford or Cambridge University, where older scholars frequently trained the younger ones (e.g., Althusser was Foucault's teacher and Foucault taught Derrida).

Frequent references to linguistics, ethnology, psychoanalysis, and Marxism are a typical feature of these philosophers. These references, however, played quite distinct roles in each case. A few thought that either conceptual analysis or philosophical intuitions could help improve specific domains, such as Marxism (according to Althusser), psychoanalysis (according to Lacan), Marxism *and* psychoanalysis (according to Deleuze and Félix Guattari or to Lyotard), or their relationships with natural science (Michel Serres). Others (e.g., Derrida) contended, rather, that disciplines such as linguistics and poetics could provide them with tools helpful for their own intrinsically philosophical concerns. Still others (Foucault and, even more radically, Deleuze) wanted to substitute a more "joyous" conception

of knowledge (cf. Friedrich Nietzsche's *The Gay Science*), free of asceticism, which has long been, according to Nietzsche, the psychological basis of the scientific spirit.

Doubtless, all these French scholars have written original works. What is more controversial is whether these works truly have been philosophically and scientifically influential or had an overall impact. They frequently and deliberately favored sophisticated rhetoric and convoluted literary phrasing, in contrast to the more typically French style with its characteristic clarity and simplicity, still exemplified at the beginning of the 20th century by the eminent French philosophers Henri Bergson (1859–1941) and Leon Brunschvicg (1869–1944) but already being challenged by Jean-Paul Sartre's work, which mixed philosophy and literature. The "collective effervescence" emerging from these often prolific scholars, however, never involved all the French philosophers interested in the social sciences. The philosopher Paul Ricoeur is a case in point, as it can be said that he was even deliberately marginalized. And far more sober works—*pace* Nietzsche—were written, which may have been more fruitful in the long term.

All these works—seductive or not—arose against an older philosophical and scientific backdrop. The rupture with Jean-Paul Sartre's existentialism, introduced by the eminent anthropologist Claude Lévi-Strauss's structuralism, played a crucial role. But Sartre's former fellows at the ENS and the Sorbonne—in particular Maurice Merleau-Ponty (1908–1961), Raymond Aron (1905–1983), Georges Canguilhem (1904–1995), and Jean Cavaillès (1903–1944)—happen to have initiated still lively trends entirely independently.

Lévi-Strauss

One can hardly understand the relationship of French contemporary philosophy with social science from the late 1950s onward if one does not mention the preeminent role of Claude Lévi-Strauss's anthropology. Although Lévi-Strauss (1908–2009) was educated—in philosophy and in law—at the same time as Jean-Paul Sartre (1905–1980), his work became influential in philosophy only after Sartre's. The latter's reputation was already at its peak immediately after World War II (Sartre's major work, *Being and Nothingness,* was published in 1943), while Lévi-Strauss's work became known to a wider audience in France only in the early 1960s. *Structural Anthropology* introduced a new paradigm in social science. Its philosophical presuppositions quickly seemed to many almost entirely opposed to Sartre's.

Structuralism in Lévi-Strauss's mind meant that the most relevant method in any social domain (from inquiries into marriage rules to the meaning of myths) is to study any element of such a domain not in isolation but always in relation to other elements and to focus on the relationships between them more than on any single elements. The role of a father in a family is variable depending on the kinds of societies in question, and similarly, other family members may also have very variable roles depending on the contextual social structures in which those roles were embedded in a distinct way in each society under study. These roles cannot be understood well without examining all close relatives' roles. When maternal uncles regularly exert authority over their nephews, as it happens in certain matrilocal societies, fathers exert a more playful role toward their children, thus inverting traditional Western roles in the same general structure exhibited by such functional relationships.

Lévi-Strauss strongly distanced himself from Sartre. He focused on the unconscious dimension of social phenomena (i.e., structures and functions as opposed to individual agents), unlike Sartre who had tried to describe the whole range of conscious mental states, especially in everyday interactions, and had strongly criticized the relevance of psychoanalysis. Lévi-Strauss did not ascribe particular significance to unconscious—possibly sexual—wishes; rather, he highlighted the unconscious social rules. He emphasized the almost deterministic strength of these unconscious rules, again in strong opposition to Sartre's emphasis on the individual's responsibility and freedom. Finally, he counterbalanced the high significance given by Sartre to history in the explanation of social facts by focusing, on the contrary, on social structures at a given time.

Lacan and Derrida

Lévi-Strauss claimed that his method had been already employed, in an exemplary fashion, by Nikolai Troubetzkoy and Roman Jacobson. These two Russian linguists admired the French-speaking Swiss linguist Ferdinand de Saussure, who had

provided linguistics with a new base (his *Cours de Linguistique Générale* was published in 1916). Saussure had anticipated structuralism with an exceptional awareness of the revolution introduced by the idea that languages should not be studied only in a historical way (as in German linguistics). On the contrary, one could display the structure of a given language at a given time (synchronic linguistics) and then compare two structures of the same language at distinct times (diachronic linguistics). Describing the progressive transformation from one structure to another structure would be a further program.

Jacques Lacan's main principle was that psychoanalysts did not have direct access to the analysands' dreams—the interpretation of which had been supposed by Freud to be the royal way to childhood's memories and to deep-down unconscious wishes—but only to the narratives of dreams. Consequently, Lacan (1901–1981) searched for interpretative tools in Jacobson's analysis of metaphors and metonymies. Furthermore, Lacan adopted Lévi-Strauss's view, according to which unconscious social rules, especially those regarding familial relationships, had to be understood in a structuralist way. This means, for example, that what plays the main formative role for a child is not necessarily his or her biological father but the one who fills the father's place in the social-familial structure.

Lévi-Strauss had not ignored the role that psychoanalysts could play in Western societies, as shamans do in other societies, and he certainly did not promote a purely abstract view of societies, in which men and women would be destined to play interchangeable roles in invariant structures. With similar concerns, Lacan thought that certain philosophers, in particular G. W. F. Hegel and Martin Heidegger, could help us understand how psychoanalysts might reveal the deep human subjecthood ("I"), based on drives (the "id") but hidden under false social identities ("ego") and irreducible to Sartre's authenticity.

Jacques Derrida (1930–2004) found specific intellectual tools in Saussure and Freud in order to refine his analysis of the presuppositions behind great philosophical works. Derrida was not searching for logical presuppositions, though; instead, his interest was in presuppositions hidden in metaphors, analogies, rhetorical questions, and the like. He called this specific method of analysis "deconstruction," which implied always viewing philosophical (and scientific) works as literary texts. According to Derrida, one of Western philosophical tradition's main presuppositions—explicit in Saussure—is the emphasis on the role of voice instead of writing (*Of Grammatology*). Derrida argued that, on the contrary, writing had extremely interesting features, in particular because it allows us to keep "traces" of the past. As a consequence, the past can be revivified as the present, in various senses. This topic is closely related to Edmund Husserl's phenomenology of time and Heidegger's ontology (*Writing and Difference*). Derrida's deconstruction of philosophical works coherently led him to explore progressively all rhetorical tools at his disposal to express his own thoughts. Doing so, he clearly went as far as the very boundaries of philosophy, as the Russian constructivist Casimir Malevich did in painting. Not surprisingly, Derrida's dialogue with John Searle, one of analytical philosophy's main representatives, was a dialogue of the deaf (*Limited Inc)*.

Althusser and Bourdieu

Lévi-Strauss also influenced Pierre Bourdieu (1930–2002), who studied philosophy almost at the same time as Derrida. First an ethnographer, then a sociologist, Bourdieu adopted Lévi-Strauss's structuralism, but unlike Lévi-Strauss, he wanted to explain how sociological and ideological structures are transmitted. He did not return to Sartre's views, however, but to the phenomenologist Maurice Merleau-Ponty's idea that usual practices and common values are often unconsciously incorporated (habitus) and so have unconscious effects. In a manner characteristic of Marxist analysis, he focused mainly on "dominating"/"dominated" relationships in social structures. Along with Jean-Claude Passeron (1930–), he adopted Gaston Bachelard's views on the necessary radical distance (break) between commonsense and scientific explanations. Later, Passeron would return to a view in favor of continuity, insisting rather on the specificity of the social sciences in comparison with the natural sciences (*Le raisonnement sociologique: L'espace non-poppérien du raisonnement naturel*, 1991).

Marxism had been renewed a little earlier by Louis Althusser (1918–1990), who distanced himself considerably from Sartre's existentialist reading of Marx. He focused, on the contrary, on the decisive break (*rupture*) constituted by *Das Kapital* in the emergence of social science, following Bachelard,

who emphasized discontinuities in the history of sciences. Althusser had a huge, although indirect, influence on the political involvement of many brilliant philosophy students during the public events of May 1968. His theoretical work, however, was almost exclusively programmatic, as the titles of his books indicate (*For Marx* and *Reading Capital*).

Foucault, Deleuze, Guattari, and Ricoeur

Michel Foucault (1926–1984), as both a historian and a philosopher, was the second French scholar after Lévi-Strauss to have influenced French philosophy greatly. The relevance and respective validity of phenomenology, psychology, and psychoanalysis had been much debated in France since the immediate postwar years. It is within this broad context that Foucault wrote his *History of Madness*.

Foucault argued that the history of the notion of madness could not be understood without being related to the emergence of modern rationality. Foucault investigated several strata of collective presuppositions, from those shared in the Middle Age—when madmen were still integrated in everyday life—to those shared in the 17th century—when madness was, according to Foucault, assimilated to absolute irrationality (an issue much debated)—and further on to those presuppositions emerging in the early 20th century—when poets like Antonin Artaud expressed in symbolism another conception of madness.

Foucault's main idea was that it made sense not to focus on the mere succession of ideas, institutions, and practices but rather to study ideas, institutions, and practices coexisting in the same period of time, even in very distant fields, and to search for their possible common hidden presuppositions. These constituted the intellectual a priori of this specific time, their "épistémè," which "archaeology" had to highlight. In *The Order of Things: An Archaeology of the Human Sciences,* he claimed that natural history in the 17th and 18th centuries had to be compared with Port Royal's general and rational grammar and Adam Smith's analysis of wealth, before being studied as the ancestor of Darwinism. Given the huge scope of his analyses, however, it has been easy to prove that Foucault's assessments were historically controversial in many cases.

Furthermore, Foucault's viewpoint was not only descriptive but also normative, and Foucault's political opinions obviously often crucially biased his assessments (e.g., on the role of asylums and prisons). Foucault's normative viewpoint was not in line with Marx's (or Freud's) but with Nietzsche's. He wanted to emphasize everything that expressed the power of life. Madness itself is sometimes correlated with creativity.

History of Madness was well received by antipsychiatrists such as David Cooper and R. D. Laing. But Foucault's work was merely historical. On the contrary, Gilles Deleuze (1925–1995), together with Félix Guattari (1930–1992), a psychiatrist and psychoanalyst initially trained by Lacan, set up a theory of schizophrenia in their joint work, *Anti-Oedipe,* published in 1972. They related this theory to a novel criticism of capitalism, inspired by Nietzsche and based on a new ontology (*Difference and Repetition*). Deleuze emphasized the relentless "flux" of life and the multiplicity of beings. Lacan, they argued, was still too concerned about the unity of subjecthood and too committed, like Lévi-Strauss, to finding invariant familial structures under cultural variations. By comparison, Deleuze and Guattari were clearly antistructuralists. They were also opposed to Hegel and to what remained Hegelian in Marx, especially dialectics as an emphasis on the permanent identity underlying temporal differences. They are sometimes viewed (by other philosophers, e.g., Vincent Descombes) as having achieved the project of synthesizing Freud and Marx more successfully than did Jean-François Lyotard (1924–1998).

Freud's theory was also examined and criticized in a more academic way by Paul Ricoeur (1913–2005) in *De l'Interpretation: Essai sur Freud,* an implicit criticism of Lacan's reading of Freud's theory as semiological. Ricoeur, in particular, pointed to the biological dimension of Freud's thought. He returned, however, to the phenomenological trend and to what had been left aside by both Lévi-Strauss and Foucault. He focused on methods in history and, more generally, on the nature of narratives, thus justifying the relevance of "hermeneutic" methods (*Time and Narrative*). Ricoeur became the doyen of this new school of hermeneutics and has been quite influential in discussions on the philosophy of the social sciences, especially in the Anglophone philosophical world. One of the main influences was the use made of Ricoeur's hermeneutics in the philosophical battle against those who wished to assimilate the social sciences to the natural sciences.

Canguilhem's, Cavaillès's, and Bachelard's Hidden Legacy

As stated above, Foucault's work was influenced not only by Nietzsche but also by the French context of distrust toward psychology, inherited in particular from August Comte and Émile Durkheim. This distrust would be challenged in French social sciences only in the 1980s by a type of anthropology congenial to experimental cognitive psychology, as practiced by Dan Sperber (1942–). Foucault's main mentor at the university was Georges Canguilhem. Canguilhem studied both philosophy and science (medicine), as did his own teacher at the Sorbonne, Gaston Bachelard (who had studied physics), and his ENS colleague Jean Cavaillès (trained in logic), who was killed by the Nazis. Canguilhem wrote numerous articles on the history of medicine, biology, and psychology, in a very sober and meticulous style. Ian Hacking (1936–), the Canadian doyen of the Stanford school of philosophy of science, while crowning his academic career as a professor in Paris (at the Collège de France), rediscovered this aspect of Foucault's heritage, which leaves relativist and Nietzschean aspects of Foucault's viewpoint entirely aside.

Gilles-Gaston Granger (1920–), who was trained by Cavaillès and influenced by Ludwig Wittgenstein and Rudolf Carnap, played a very significant role in the French philosophy of social science by distancing himself from Marx quite early on. Although admiring Lévi-Strauss for his use of formal models in the social sciences, he broadened attention to other domains of social science, such as Keynesian economics and Decision Theory (*Pensée Formelle et Sciences de l'Homme*). A generation later, Jacques Bouveresse (1940–), in line with Wittgenstein as well, sharply criticized the rhetorical abuses of certain French philosophers and sociologists (such as Bruno Latour), as well as their relativism.

Michel Serres (1930–), who was also trained in the same intellectual climate, was, however, more specifically influenced by Gaston Bachelard. Bachelard was interested not only in science but also in the study of myths as expressing another dimension of human beings. Serres brilliantly illustrated conceptions close to Foucault's episteme (*Hermès*). To some extent, he went further than Foucault on this path, by arguing that natural science, such as thermodynamics, should be studied jointly with literature—for example, Jules Verne's novels—because they might express the same implicit presuppositions during a certain historical period. His way of writing, however, is not analytic, conceptually speaking, but very literary, as are Lacan's, Derrida's, Foucault's, Deleuze's, and, to some extent, Bourdieu's. He claimed to be searching for bridges from social sciences to natural sciences (*Le Passage du Nord-Ouest,* 1980) but tended to comment poetically on this passage more than to construct real bridges.

Aron's Paradoxical Legacy

In the long term, Raymond Aron might have had the deepest influence on contemporary French philosophy of social science. Politically oriented to the right, he nevertheless admired Marx's work. For many, he turned out retrospectively to have often made more accurate assessments than Sartre regarding the significance of both political events and emerging ideas. As he was more open to ideas coming from the English-speaking world compared with other French scholars, left-wing as well as right-wing PhD students found inspiration in his work. Typically, the Norwegian Marxist social scientist Jon Elster (1940–) preferred to write his PhD thesis—on Marx—under Aron's supervision rather than under Althusser's.

Aron also reintroduced the work of Max Weber in France and rediscovered Alexis de Tocqueville, who had been forgotten during the dominance of Marxism in France. Within the same tradition, Raymond Boudon (1934–), also educated as a philosopher at the ENS, introduced mathematical modeling into the practice of French sociology. He has also conducted an ongoing critical analysis of the fundamental concepts and logical presuppositions of contemporary social sciences (including a sharp criticism of structuralism), based on an in-depth knowledge of these disciplines.

Paul Veyne (1930–), formerly a regular attendee at Foucault's seminars, is both an outstanding historian and a subtle philosopher of social science. He has recognized that he shares Aron's views on history. Like Boudon's (and Elster's), this kind of philosophy of social science is intimately connected with the practice of the social sciences.

Conclusion

Comparable with German philosophy, yet unlike American philosophy, French philosophy has often been closely connected to the social and human sciences. Ethnology and psychoanalysis, in particular,

have long captured the attention of French philosophers. These two disciplines seemed to blaze a third trail between natural science and a social science inspired by natural science, such as economics. But unlike German philosophy, the so-called French theory has often deliberately tried to mix philosophy with literature, sometimes successfully if considered in accordance with aesthetic canons. The way was thus opened to a French vogue that swept many departments of literature in universities in the United States. French philosophy of social science, however, is much broader, encompassing works that abide strictly by epistemological rules, very much in line with the prevailing French tradition of René Descartes, Nicolas de Condorcet, and Antoine Cournot.

Alban Bouvier

See also Durkheim's Philosophy of Social Science; Foucault's Thought; Habitus; Marxism and Social/Historical Explanation; Postmodernism; Psychoanalysis, Philosophical Issues in; Relativism and the Social Sciences: From the Sapir-Whorf Hypothesis to Peter Winch; Structuralism and Poststructuralism; Unconscious

Further Readings

Althusser, L. (1969). *For Marx*. London, England: New Left Books. (Original work published 1965)

Althusser, L., & Balibar, E. (1971). *Reading capital*. New York, NY: Pantheon. (Original work published 1965)

Aron, R. (1948). *Introduction to the philosophy of history: An essay on the limits of historical objectivity*. London, England: Weidenfeld & Nicolson.

Aron, R. (1957). *The opium of the intellectuals*. London, England: Secker & Warburg.

Bachelard, G. (1938). *The psychoanalysis of fire*. Boston, MA: Beacon Press.

Boudon, R. (1986). *Theories of social change: A critical appraisal*. London, England: Basil Blackwell.

Canguilhem, G. (1988). *Ideology and rationality in the history of the life sciences*. Cambridge: MIT Press.

Cusset, F. (2008). *French theory: How Foucault, Derrida, Deleuze, & Co. transformed the intellectual life of the United States*. Minneapolis: University of Minnesota Press.

Deleuze, G. (1994). *Difference and repetition*. London, England: Athlone.

Deleuze, G., & Guattari, F. (1984). *Anti-Oedipus: Capitalism and schizophrenia* (R. Hurley, M. Steem, & H. R. Lane, Trans.). London, England: Athlone.

Derrida, J. (1976). *Of grammatology*. Baltimore, MD: Johns Hopkins University Press.

Derrida, J. (1978). *Writing and difference*. Chicago, IL: University of Chicago Press.

Derrida, J. (1988). *Limited Inc*. Evanston, IL: Northwestern University Press.

Descombes, V. (1980). *Modern French philosophy*. Cambridge, England: Cambridge University Press.

Foucault, M. (1970). *The order of things: An archaeology of the human sciences*. London, England: Tavistock.

Foucault, M. (2006). *History of madness*. Abingdon, England: Routledge. (Original work published 1961)

Habermas, J. (1990). *The philosophical discourse of modernity*. Cambridge: MIT Press.

Lacan, J. (1977). *Écrits*. New York, NY: W. W. Norton.

Lévi-Strauss, C. (1968). *Structural anthropology*. London, England: Penguin.

Ricoeur, P. (1977). *Freud and philosophy: An essay on interpretation*. New Haven, CT: Yale University Press.

Ricoeur, P. (1984–1985). *Time and narrative* (Vols. 1–3). Chicago, IL: University of Chicago Press.

Serres, M. (1968). *Hermes: Literature, science, philosophy*. Baltimore, MD: Johns Hopkins University Press.

Sartre, J.-P. (2004). *Critique of dialectical reason*. London, England: Verso. (Original work published 1960)

Veyne, P. (1984). *Writing history: Essay on epistemology*. Middletown, CT: Wesleyan University Press.

Context of Discovery Versus Context of Justification

The basic assumption behind the discovery/justification (D-J) or "two-context" distinction, as employed by Hans Reichenbach and the Logical Empiricists—but also by Karl Popper—is that there exists a general logic of justification in science but no *logic* of discovery. This entry explains the aim of the upholders of this distinction (once considered important for genuine scientific method and thus shaping official philosophy of science for a certain period), reviews criticisms against it, and assesses some positive uses of it.

Aim

For these methodologists, creative acts, such as inventing hypotheses or composing piano sonatas, are idiosyncratic and to be left to psychologists and historians to study. Since these philosophers

conceived scientific method as the logic of science, and since there did seem to be a routine method for testing hypotheses against data, one analogous to the systematic method of checking logical proofs, they employed the distinction invidiously to *restrict* scientific method and epistemology to the logical structure of theories, predictions, explanations, and so on, already on the table. Accordingly, these philosophers rejected the foundationist views of Bacon, Descartes, and Newton, according to which using the correct method of discovery itself is the strongest form of justification. Since the Logical Empiricists and Popper shaped modern, academic philosophy of science, the D-J distinction became entrenched, leaving creative activity as an *exogenous* factor in philosophical models of science.

Criticisms

The D-J distinction and its applications invited several criticisms.

1. The distinction is unclear. Its typical formulations harbor several distinctions that do not stand or fall together. For example, if the context of justification is understood as a temporal phase of research rather than as a timeless logical relation of theoretical to observation statements, then the context of justification is filled with discovery tasks such as searching for predictions that the theory implies and inventing the experimental means of checking them. Artificial intelligence (AI) experts were the first to emphasize this point, noting that even the discovery of a deductive logical proof is an inductive search task—a discovery task—one that needs to be guided by heuristics to be efficient.

2. By switching the emphasis from romantic notions of theory formation to problem solving as the central activity of scientists, AI experts made the search for systematic discovery methods a more reasonable enterprise. For Herbert Simon and colleagues, discovery is problem solving, and problem solving is a search through a space of possible solutions. Hence, discovery is ubiquitous in research. There cannot exist a pure context of justification in the temporal sense.

3. Logical empiricist attempts to formulate a theory of confirmation in logical or probabilistic terms were failures. Besides generating paradoxes, an untenable distinction between a pure observation language and a theoretical language undercut these efforts.

4. Norwood Russell Hanson retained the D-J distinction but asserted the possibility of a logic of discovery. He revived Charles S. Peirce's topics of abductive logic of discovery and economy of research and posited an intermediate stage of preliminary evaluation between the contexts of discovery and justification, an idea developed by later writers under labels such as "prior appraisal," "pursuit," and "heuristic appraisal."

5. By the 1960s, most philosophers were abandoning logical empiricism. Historically oriented analysts such as Thomas Kuhn, Paul Feyerabend, Stephen Toulmin, and Imre Lakatos denied that there is a systematic logic of justification any more than there is a logic of discovery and insisted that the products of science cannot be fully understood in isolation from the process that produced them.

6. Kuhn often took the viewpoint of research scientists who constantly engage new problems at research frontiers, a pro-discovery stance excluded by the positivists' use of the D-J distinction.

7. Henceforth, philosophers paid more attention to creativity and invention in the history of science and drew on the resources of AI and other cognitive sciences, as well as the heuristic power of rhetorical tropes (analogy, similarity, metaphor, and resemblance) and their application in modeling. Some turned upside down the positivist view of heuristics as mere temporary scaffolding, regarding such guides as fundamental to the advance of knowledge.

8. Although the new social studies of science avoids the term *discovery,* given its apparent commitment to theoretical realism (the view that the sciences are approaching the final truth about reality), much work in this field pertains to "science in action" and hence to "discovery" in the broad, pragmatic sense of scientists' invention or social construction practices at research frontiers.

Positive Implications

Despite these largely negative responses to the invidious use of the D-J distinction since the 1960s, some applications of it remain useful. General confirmation theory still thrives in the form of subjective Bayesian probabilistic approaches. Evolutionary epistemology

also motivates a D-J distinction. Donald Campbell has argued that all innovation, all inductive increases of fit, must be the product of undirected variation plus selective retention, a generalized Darwinian selectionist process defended by Popper as well. For at the frontier of knowledge, scientists can proceed only by trial and error—by a variation–selection process. While the variant ideas and practices produced in the search for new problems and solutions are always constrained by current knowledge to some degree, thereby cutting down the search space, the remaining space can be explored only by trial and error. This remaining element of blind guessing falls within the context of discovery, while the testing phase falls within the context of justification. AI experts have devoted much attention to both contexts, and to pursuit as well, in the form of economy of research—the study of which search strategies are best for problems of a given kind. Indeed, the field of evolutionary computation, including genetic algorithms, has turned selectionism into a powerful approach to automated discovery, thus reviving to some degree the old idea of methodology of discovery.

Thomas Nickles

See also Abduction and Inference to the Best Explanation; Artificial Intelligence; Induction and Confirmation; Logical Positivism/Logical Empiricism; Popper's Philosophy of Science; Scientific Method; Verificationism

Further Readings

Creview, D. (1993). *AI: The tumultuous history of the search for artificial intelligence.* New York, NY: Basic Books.

Hanson, N. R. (1971). The idea of a logic of discovery. In N. R. Hanson (Ed.), *What I do not believe and other essays* (pp. 288–300). Dordrecht, Netherlands: D. Reidel.

Koza, J. (1992). *Genetic programming: On the programming of computers by means of natural selection.* Cambridge: MIT Press.

Nickles, T. (Ed.). (1980). *Scientific discovery, logic, and rationality.* Dordrecht, Netherlands: D. Reidel.

Reichenbach, H. (1938). *Experience and prediction.* Chicago, IL: University of Chicago Press.

Schickore, J., & Steinle, F. (2006). *Revisiting discovery and justification: Historical and philosophical perspectives on the context distinction.* Dordrecht, Netherlands: Springer.

Conventions, Logic of

This entry traces the conceptual and logical origins of convention, a ubiquitous social phenomenon, by presenting four main theoretical accounts of it.

One of the earliest attempts to delineate a role for conventions in social analysis was David Hume's influential discussion of justice in *A Treatise of Human Nature*. His ideas have resurfaced in theoretical contributions to economics, sociology, and law, as well as being the inspiration for David Lewis's seminal philosophical study of convention. As this entry will show, Hume's ideas are fertile, and subsequent analyses that claim to develop them have focused on different elements. However, a common theme that most modern approaches share is a preoccupation with understanding *social rules* that are both *normative* (in the sense that we should follow them) and *arbitrary* (in the sense that they might have been otherwise).

Classic examples of social conventions include the convention of driving on the right in some countries but not in others, various rules of etiquette and dress, and also language, fiat money, and the foundations of law (at least for legal positivism).

Hume on the Origins of Justice

Hume tells us that justice—which he understands very minimally as a system of property rights—is an "artificial virtue" because it cannot be established by the unhampered interaction of the natural passions. Yet systems of property—while nonnatural in the sense that they do not flow directly from the operation of innate dispositions—exist. According to Hume, this is made possible by a human "artifice" or *convention*. The artifice is fortuitous, because the narrow self-interest of those who manage to move from an unstable, impoverished state of nature to a society with rules of property ownership tends to be well served. Individuals can engage in a collaborative project that extends beyond the family and allows them to enjoy the fruits of their labor safely.

Hume defines conventions as "a general sense of common interest" expressed by all the members of society to one another. This shared sense of common interest motivates them into regulating their conduct in accordance with certain rules. Crucially, conventions depend on reciprocity and mutual recognition

for their force and thus can exist prior to contracts or promises. As Hume (1978) notes,

> This may properly enough be call'd a convention or agreement betwixt us, tho' without the interposition of a promise; since the actions of each of us have reference to those of the other and are perform'd upon the supposition, that something is to be perform'd on the other part. (p. 490)

In Hume's state of nature, primitive conventions are the primary mechanism for neutralizing the potentially conflicting interests of nonsocialized individuals. Conventions come about in situations where a number of people each stand to gain by adopting a principle that regulates interaction between them. In the case of property rights, Hume describes a principle of reciprocity: Individuals agree not to take the goods that another has rightfully gained, provided that their own goods will not be taken in a similar manner. However, once the principle is transgressed, the system breaks down because it is not underwritten by moral obligations. He also assumes that the adoption of principles such as these must involve some linguistic interaction: People must be aware that mutual interdependence exists. This communicative element is essential because there is no preexisting framework of obligation, right, or duty—without signaling, it is unclear why primordial men would be motivated to take the risk of being exposed to transgressions.

Although he was primarily concerned with justice, Hume suggested that language, money, and a host of other social institutions are conventions. In order to explain these, he introduced a second version of convention, which is exemplified by the example of rowers seamlessly *coordinating* in order to propel a boat across a river. Here, Hume introduced a new intuition about conventions—that they can arise gradually by small adjustments in behavior prompted by a *shared goal.* The rowers are not in the same situation as primordial men because they have a common, unique, and immediate objective—to move the boat through the water. We can easily imagine a situation in which given a number of possible rowing techniques, they would be completely indifferent between them. Moreover, they can coordinate their behavior without declaring interests because their mutual desire for coordination is trivial. In this situation, principles no longer seem to play a role—conventions can be established by channeling expectations, a simple regard for precedent, and a method for identifying transgressions.

While the first version stresses the mutual expression of interdependence of *selfish* interests, Hume's second version of convention shows how conventions arise out of a type of *inductive,* trial-and-error-based reasoning where language is not essential. It was a development of the latter interpretation by philosophers that has generated the greatest volume of philosophical and social-scientific research on conventions in recent years.

David Lewis and Game Theory

David Lewis, the pioneer of the game-theoretic approach to convention, acknowledged his debt to Hume by suggesting that the formal structure that he elaborated was "merely scaffolding" that restated earlier Humean ideas. Nevertheless, it is his formalization of earlier intuitions that has held sway in recent debates in both philosophy and the social sciences.

Lewis sees conventions as arising from situations where agents have a mutual coincidence of wants, and he analyzes these situations as *coordination games.* He defines a coordination game as an interdependent decision by two or more agents where there is a coincidence of preferences and two or more "proper" coordination equilibria. A coordination equilibrium strategy is one where players would not have wanted to act otherwise after the outcome has been reached. For game theorists in the Lewisian mold, potential participants in a convention are faced with the following problem: How should they coordinate around one of a number of equally satisfactory possible outcomes given that any coordinated behavior is preferable to a failure to coordinate?

Their answer relies heavily on the importance of *precedent.* If information about past iterations of the game is available to players, then they would have a reason to coordinate by employing a "salient" strategy that has been employed in the past. The past equilibrium provides them with a focal point for future coordination. Lewis (1969) summarizes the essence of the game-theoretic conception as "a self-perpetuating system of preferences, expectations, and actions capable of persisting indefinitely" (p. 42). Precedent thus leads to the formation of

expectations about future behavior, which, in turn, leads to *conformity* in the actions of individuals. This system of expectations and actions is represented by a behavioral regularity that conforms to one of the possible equilibrium strategies of the game.

The only way for agents to conform to a convention is through calculations based on their expectations of other people's behavior. However, this can lead to an infinite regress of higher-order expectations, which would normally be resolved by a contract or promise. It is because (following Hume) game theorists reject this option that convention becomes dependent on precedent. Yet the theory needs a further "common knowledge" assumption in order to avoid infinite regress. Lewis tells us that x is common knowledge in a population y if every member of that population knows x and knows that everyone else knows it, and so on. His model thus requires that for a given population to sustain a convention, it must be common knowledge that a particular behavioral regularity is the convention.

More recently, the game-theoretic rendering of convention has gained in mathematical sophistication and scope within economics, and evolutionary games have supplanted classical games. In this context, agents are assumed to be uniquely backward looking, basing their decisions solely on the observation of past behavior. Agents are also reactive, following precedent unthinkingly while remaining unaware of potential alternatives and unable to effect change through deliberation. The revised structure allows modelers to "predict" which out of a number of competing behavioral regularities will emerge as the convention. Though "deviant" behavior is possible, the system as a whole will converge on a unique, long-run equilibrium. Evolutionary games account for the emergence of conventions, without relying on standard rationality assumptions or the coincidence of preferences. Instead, conventional behavior is modeled as a *self-organizing system* that is independent of its own history.

Agreement Theory

Lewis's aim was to show how rational agents could follow conventions *without prior agreement*—purely on the basis of precedent. However, critics of the game-theoretic approach have identified a tension in his analysis. They claim that game theorists wishing to adhere to a classical (economic) conception of rationality cannot accept a Lewisian account of coordination through precedent. This is because classical rationality is exclusively forward looking: A strategy is rational at time t if and only if it maximizes expected utility from t into the indefinite future. Precedent might allow agents to coordinate expectations, but once the rationality of agents is common knowledge in a given population, expectations must be based on the canons of rationality rather than on the reproduction of past behavior. For a rational convention follower, past conformity cannot be an adequate reason for present conformity, nor would she expect another rational convention follower to conform in the present just because she has conformed in the past. Common knowledge of precedent will not, therefore, suffice to generate either expectations of conformity or actual conformity.

This argument (among others) has led one of Lewis's critics, Margaret Gilbert, to develop an alternative conception of convention. In doing so, she abandons one of the basic assumptions of game theory—the idea that conventions can be established without prior agreement or communication. The result is what we might call *an agreement theory of convention,* which emphasizes the *quasi-moral* features of coordination. Gilbert observes that conventions, despite their informal and ephemeral nature, are ubiquitous and influential in everyday life. Though they are rarely consciously followed and lack the rigidity of moral prescriptions, the conventions of a community are usually described in moral language, which permeates the analysis of the concept. Like agreements, conventions have positive and negative implications once adopted—we feel that we "ought" to abide by the prevailing conventions and that we "ought" to avoid behavior that violates them. Our attitudes are not exclusively based on individual cost–benefit calculations. Harking back to Hume's first version of convention, agreement theory explains this quasi-moral aspect of conventions by claiming that they are conditional on the tacit joint acceptance of principles of action. These principles take a fiat form—they are stipulated by participants in the conventional practice rather than justified by reference to a higher-order normative framework, such as the common good or self-interest. Thus, agreement theory replaces assumptions about individual calculative rationality

with the myth of an original contract and explains coordination through this *contractarian* device.

A Keynesian Alternative

We have now seen how two divergent themes that emerge from the Humean account of convention have been reinterpreted and transformed in recent philosophical and theoretical work. A third influential approach is the Keynesian tradition within economics, which also treats convention as a basic theoretical tool, though it eschews the Humean framework. This should not come as a surprise, because Keynes's interest in convention was derivative of his desire to understand an economy characterized by radical *uncertainty*. Keynes believed that investment decisions (and many other economic choices) were made under conditions of radical uncertainty. That is to say, he thought that calculating the probability of future states of the world and assigning quantifiable risks to them was not possible. Furthermore, Keynes argued that uncertainty was not only an inevitable feature of economic psychology, it was also ontologically grounded—the economy is a nonergodic system (i.e., a system where the statistical rates true of the ensemble of actors do not correspond with, or are not equal to, statistics true of a given individual actor, e.g., an investor).

According to Keynes, under true uncertainty the only way rational agents might avoid paralysis is to follow a convention. In order to illustrate this, he provided a typology of conventions that, he claimed, operate in the financial markets. They include the assumptions that (a) the present is a "serviceable guide" to the future, (b) the existing state of opinion (as expressed by prices) accurately sums up future prospects, and (c) individuals attempt to conform to the behavior of the majority. Thus, Keynes treated conventions as *coping mechanisms* used by fallible agents in volatile and unpredictable socioeconomic environments. Subsequent Keynesian analysis has developed this theme by conceptualizing conventions as stabilizing particular, contingent social arrangements without introducing (neoclassical) economic notions of equilibrium. In recent years, this approach has been championed by proponents of the French *économie des conventions* school, with particular emphasis on the forms of coordination that result from the operation of radical uncertainty.

John Latsis

See also Coalition Logic; Collective Goals; Cooperation/Coordination; Evolutionary Game Theory and Sociality; Promises and Agreements; Social Contract Theories; Social Conventions; Social Norms; Social Rules

Further Readings

Gilbert, M. (1996). *Living together: Rationality, sociality, and obligation*. Lanham, MD: Rowman & Littlefield.

Hume, D. (1978). *A treatise of human nature*. Oxford, England: Clarendon Press. (Original work published 1739)

Keynes, J. M. (1937). The general theory of employment, interest and money. *Quarterly Journal of Economics, 51*(2), 209–223.

Latsis, J., De Larquier, G., & Bessis, F. (2010). Are conventions solutions to uncertainty? Contrasting visions of social coordination. *Journal of Post Keynesian Economics, 32*(4), 535–558.

Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Cooperation, Cultural Evolution of

This entry presents an account of the indispensable social phenomenon of cooperation from the viewpoint of modern theories of evolution. In particular, the entry sketches the outline of a process in which cultural evolution, or the transmission of behaviors and beliefs among individuals via social learning, can explain the full range of human cooperation.

Cooperative Behavior and Kin Selection

Cooperative acts, those that provide a benefit to others at a cost to the individual performing them, are common across all cultures. Consider food sharing in hunter-gatherer societies, voting in modern societies, or fighters willingly dying in battle for their tribe or country. From an evolutionary perspective, that any individuals cooperate at all is puzzling, because we expect evolution by natural selection to result in behaviors that benefit the individual performing them. Human cooperation, furthermore, is particularly vexing, because unlike in other organisms, in human societies, cooperation involves hundreds, thousands, or even millions of *unrelated* individuals engaging in *unrepeated* interactions. The common

explanations for cooperative behavior in other organisms are kin selection and reciprocity. Kin selection involves the evolution of a genetic propensity to provide benefits to related individuals because they are more likely than average to share your cooperative genetic heritage. Reciprocity involves the evolution of a propensity to provide benefits to others, regardless of relationship, if they are likely to return the favor in the future. Kin selection and reciprocity have undoubtedly played an important role in the evolution of human cooperation; nepotism and reciprocal favors are ubiquitous. Still, these two mechanisms alone are unable to explain cooperation among unrelated individuals in unrepeated encounters.

Learning

Social learning, or the acquisition of behaviors and beliefs from another individual through teaching or direct imitation, is one way an individual can acquire useful information about the environment. When information is costly and environments are variable, it can be adaptive to rely on social learning to shape behavior (rather than learning individually by trial and error, or using a relatively inflexible set of innate behaviors). Because it is not always clear whom to learn from in a complex social environment, social learning can often take the form of simple general strategies, such as "copy the majority" or "copy the most successful other individual." Such strategies allow individuals to efficiently acquire appropriate behavior in a complicated world and thus are a likely product of evolution by natural selection under a range of conditions. Over longer periods of time, transmission of information among individuals via social learning results in a process of cultural evolution in the sense that behaviors become more or less common in the population based on whether or not they are transmitted successfully. As individuals reshape and retransmit information they acquire from others, cultural complexity accumulates such that no single individual, in his or her own lifetime, can reinvent the complete body of knowledge stored in the entire population. Cumulative cultural evolution, a human hallmark, is therefore an important emergent property of the evolution of the capacity to learn socially.

Norms and "Cultural Group Selection"

One characteristic of cumulative cultural evolution particularly relevant to cooperation is that it has the capacity to create different norms, or stable, shared behavioral standards, among different groups of individuals. For example, if many individuals follow the behavior of the single most successful or prestigious founder of a society, then the group at large will tend to develop a specific norm. If individuals also have a tendency to copy the majority, then a norm can be maintained within a group, even if new people or ideas migrate into the group at high rates, because new arrivals will quickly conform to the preexisting norm in the group. Via such processes, different groups will evolve different norms of behavior; some groups will be more cooperative, while other groups will be less cooperative. The emergence of variation in cooperative norms among groups establishes the conditions for intergroup competition, in which norms from successful groups spread to other, less successful groups. This process, known as "cultural group selection," can occur when cooperative groups outcompete other groups in direct conflict or war, when they produce more individuals who subsequently migrate to other groups, or when they are preferentially imitated by individuals in less successful groups. While genetic group selection is generally thought to be an unimportant evolutionary process in genetically well-mixed species, such as humans, cultural group selection is plausible because the culture-specific evolutionary processes outlined above work to maintain cultural variation *between* groups while reducing cultural variation *within* groups.

Stable Environments and "Gene–Culture Coevolution"

The cultural evolution of cooperation as described above thus results in the creation of novel and stable environments, which can then influence the further genetic evolution of cooperative tendencies. In an environment of stable cooperation, genes that predispose individuals to identify and follow local, cooperative norms will be favored. For example, moral norms lead cooperators to collaborate to reduce the genetic fitness of norm violators. This can lead to the evolution of innate propensities to follow norms, as well as impulses, such as guilt and shame, that compel people to behave cooperatively. Ultimately, then, human cooperation is the result of a complicated coevolutionary process of selection on cultural variation and selection on genes. In summary, the evolution of a capacity for social learning leads to

cumulative cultural evolution and the emergence of stable, between-group behavioral variation. This variation then potentiates cultural group selection and the spread of cooperative norms across groups, which in turn leads to the evolution of innate psychological tendencies adapted to a world of cooperative norms. This "gene–culture coevolutionary" process complements our understanding of the evolution of cooperation via kin selection and reciprocity and can help explain how and why humans are unique in cooperating in large, ephemeral groups of unrelated individuals.

Peter J. Richerson and Vicken Hillis

See also Biology and the Social Sciences; Cooperation/Coordination; Cultural Evolution; Evolutionary Psychology; Human Cultural Niche Construction and the Social Sciences; Social Norms; Sociobiology

Further Readings

Bowles, S., & Gintis, H. (2011). *A cooperative species: Human reciprocity and its evolution*. Princeton, NJ: Princeton University Press.

Chudek, M., & Henrich, J. (2011). Culture-gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences, 15*(5), 218–226.

Richerson, P., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago, IL: University of Chicago Press.

Cooperation/Coordination

Adam Smith pointed out in *The Wealth of Nations* that man is unique among God's creatures in that he stands at all times in need of the cooperation of a great multitude of his fellow men for his very survival. Nature is red in tooth and claw, and the human being is not physically well provided to survive. So we must cooperate with our fellow humans. But such cooperation is usually the realm of intimate friends and those in kinship relationships with us. The mystery of modern life is that the individual has the time and opportunity to make but a few close personal friends in his or her lifetime but requires the cooperation of hundreds, thousands, and even millions of others to be able to survive.

Smith's reasoning was that since one is relying on the cooperation of others but these others are unknown to one and perhaps will always remain unknown, the puzzle is how to achieve cooperation among strangers. He argued that one cannot rely on an appeal to their benevolence but instead must make an appeal to their self-interest. It is not from the benevolence of the butcher, the baker, and the brewer that we secure our dinner but through their self-interest.

In marshaling the self-interest of some to meet the demands of others, social cooperation under the division of labor is achieved. But Smith, in passages in the first book of *The Wealth of Nations,* delves into the complex coordination of economic activities that is required to realize the gains from trade. As he puts it at one point, the number of exchanges would exceed all computation, yet it is through the market mechanism that the complex coordination of the activities that constitute the division of labor in any society is achieved. Cooperation is realized because coordination is achieved.

This is the most important meaning of the terms *cooperation* and *coordination* in economics as historically contemplated. But in more contemporary treatments, these terms often take on different meanings—usually associated with behavior within a model. For example, businessmen might cooperate with one another to set the price in a model about cartel behavior. Or in a model of market equilibrium the production plans of some are meshed with the consumption demands of others, such that a full coordination of economic plans occurs. Cooperation and coordination are important attributes of a variety of game-theoretic models in the social sciences. The question remains as to how the attribute of the model dovetails with the more real-world notion of cooperation and coordination that Smith was talking about.

A canonical game-theoretic model is that of the Prisoner's Dilemma. In that model, the possibility of cooperation among the players is precluded by design, and each is compelled by the logic of the situation to choose a suboptimal path of defect–defect, rather than the collectively preferred option of cooperate–cooperate. This model has been utilized to illuminate a surprising number of complex problem situations in the real world—from the instability of cartel agreements to the escalating arms race during the Cold War.

A more recent vein of research tried to demonstrate how through small adjustments social cooperation could actually be achieved even in similar Prisoner's Dilemma situations. The possibility of cooperation arising out of situations of conflict has tended to highlight various mechanisms of reputation, and other punishment strategies that shift the payoffs that agents face when they don't cooperate and instead cheat. Robert Axelrod's *The Evolution of Cooperation* demonstrated that in iterative interactions, players benefited more from cooperation than from cheating, even in situations such as a one-shot game, where the rational strategy would have been to defect. In experimental economics, as Vernon Smith has shown, the results of the vast majority of experiments reveal that more cooperation is possible than the standard model would predict.

The existence of common-pool resources has also raised issues related to the interests of individuals and groups. A common-pool resource is defined by two fundamental characteristics: (1) one person's use of a unit of common-pool resource makes that same unit unavailable to anybody else and (2) it is costly to exclude potential users of a common-pool resource. Some classic examples of common-pool resources include fisheries, forests, underwater basins, and grazing lands. In a 1968 article, "The Tragedy of the Commons," Garrett Hardin argued that because of the unique characteristic of the commons, its users were in a situation that ultimately led to the destruction of the very resources they depended on—when, for example, individuals overfished or overgrazed resources for immediate individual gain. To avoid the tragedy from playing out, Hardin suggested that the commons should be either privatized or managed by the government, so that the right to entry and use could be controlled. However, Elinor Ostrom showed through a variety of field case studies across time and place that cooperation is indeed possible as communities evolve rules of self-governance to protect common-pool resources. In the Ostrom study, as long as rules come to limit access, assign accountability for use, and introduce graduated penalties for inappropriate use of the resource, cooperation through self-governance is indeed possible.

In fact, one way to reconcile the insights of Scottish Enlightenment thinkers such as Smith on cooperation and coordination, and the modern game-theoretic literature in political economy is to draw inspiration from Ostrom's work on institutional analysis. The self-interest postulate is squared with the invisible-hand postulate through institutional analysis. Or to put it another way, Ostrom has provided a multitude of evidence from the field and in the lab to show that out of a Hobbesian problem situation (after Thomas Hobbes's description of the state of nature as a war of all against all) a Smithian solution concerning social cooperation under the division of labor can be achieved provided the rules of governance are in operation.

Peter J. Boettke

See also Collective Rationality; Cooperation, Cultural Evolution of; Evolutionary Game Theory and Sociality; Invisible Hand Explanations; Markets and Economic Theory; Scottish Enlightenment, Influence on Social Sciences; Social Contract Theories; Social Conventions; Sociobiology

Further Readings

Axelrod, R. (1985). *The evolution of cooperation.* New York, NY: Basic Books.

Boettke, P. (2001). *Calculation and coordination: Essays on socialism and transitional political economy.* London, England: Routledge.

Hardin, G. (1968). The tragedy of the commons. *Science, 162,* 1243–1248.

Ostrom, E. (1990). *Governing the commons.* New York, NY: Cambridge University Press.

Seabright, P. (2003). *The company of strangers.* Princeton, NJ: Princeton University Press.

Smith, A. (1977). *An inquiry into the nature and causes of the wealth of nations.* Indianapolis, IN: Liberty Fund. (Original work published 1776)

Smith, V. (2007). *Rationality in economics.* New York, NY: Cambridge University Press.

Cost–Benefit Analysis

Cost–benefit reasoning has been a central idea not only in economics but also in several social-scientific disciplines and in philosophy as well, especially in certain approaches to moral and political theory. Reasoning according to a cost–benefit analysis (CBA) is seen as a crucial element of rationality in action and is therefore important for any philosophical understanding of social action.

This entry introduces the notion, shows the key role it plays in policy making, and gives a critical overview of the various approaches on offer.

Overview

CBA is a tool for evaluating governmental policies. It is widely employed by applied economists and, increasingly, by governments. CBA ranks policies by summing willingness-to-pay/accept (WTP/WTA) amounts (as estimated from behavioral evidence or surveys). Consider two policies, *a* and *b*. For simplicity, assume that *a* leads for sure to outcome *x* and that *b* leads for sure to outcome *y*. Imagine that Joe prefers *y* to *x*. Joe's WTP for *y*, taking *x* as baseline, is the reduction in his monetary income in *y* that just suffices to make him indifferent between the outcomes. Similarly, if Sue prefers *x* to *y*, then her WTA for *y*, taking *x* as baseline, is the increase in her income in *y* that just suffices to make her indifferent between the outcomes.

CBA then compares policies *a* and *b* by seeing whether *x* (the outcome of *a*) has net positive monetized benefits as compared with *y* (the outcome of *b*). Take the aggregate WTP amounts of the individuals who prefer *y*, and subtract the aggregate WTA amounts of the individuals who prefer *x*. Call this value for short the "WTP/WTA aggregate" for *y* relative to *x*. CBA ranks *a* over *b* just in case this value is positive.

The definition of CBA just provided smoothly generalizes to the more realistic case where policy outcomes are uncertain—in that case, each policy corresponds to a probability distribution over outcomes—but to ease presentation, this entry will focus on the simpler case and will frame the discussion directly in terms of outcomes rather than the policies giving rise to them.

Approaches and Critiques

CBA has been criticized on numerous grounds, and the most important are the following.

Ordering Failure

CBA can actually fail to rank outcomes in a minimally rational manner. It turns out to be possible that the WTP/WTA aggregate for *y*, relative to *x*, is positive (so that CBA says *y* is a better outcome) but that the WTP/WTA aggregate for *x*, relative to *y*, is also positive. It is also possible that everyone prefers *z* to *y* but that the WTP/WTA aggregate for *z*, relative to *x*, is less than the WTP/WTA aggregate for *y*, relative to *x*.

These ordering failures, however, can be solved via a refinement of the CBA test. Given a set of outcomes, arbitrarily choose one (*x*) as baseline. For every other outcome *y*, assign it an overall amount equaling the *negative* WTP/WTA aggregate for *x*, relative to *y*. (Assign *x* an overall amount of 0.) Rank outcomes according to these overall amounts. This ranking will be transitive; there will never be reversals, with each of a pair of outcomes ranked better than the other; and if everyone prefers one outcome, it will be ranked better.

Kaldor-Hicks Efficiency and the Justification of CBA

One outcome is *Pareto superior* to a second if everyone prefers the first. The CBA test does not, of course, guarantee Pareto superiority. The WTP/WTA aggregate for *x*, relative to *y*, can be positive, even though some prefer *x* and others, *y*.

However, CBA is often defended by invoking the idea of *potential* Pareto superiority or, equivalently, Kaldor-Hicks efficiency. Outcome *x* is Kaldor-Hicks efficient, relative to *y*, if there is a hypothetical, costless redistribution of resources—converting *x* into *x** such that *x** is Pareto superior to *y*. Leaving aside some technical issues, it is roughly true that if the WTP/WTA aggregate for *x*, relative to *y*, is positive, then *x* is Kaldor-Hicks efficient relative to *y*.

Less formally, one outcome passes the CBA test, relative to a second, if those who gain from the first outcome could, in principle, fully compensate those who lose, via a scheme of compensation payments with no administrative costs—leaving everyone better off.

However, the nexus between CBA and *hypothetical* compensation does not really furnish much of a justification. If *x* is Kaldor-Hicks efficient relative to *y*, but not Pareto superior, there are some individuals who prefer *y*. Why do the premises that (a) everyone prefers *x** to *y* and (b) *x* is transformable into *x** warrant the conclusion that (c) *x* (and not just *x**) is better than *y*? To tell those made worse off by some governmental policy that they *could* be compensated hardly answers their complaints about the policy if compensation is not, in fact, forthcoming.

A better justification for CBA sees it as a rough proxy for overall well-being. The fact that the WTP/

WTA aggregate for *x*, relative to *y*, is positive indicates (albeit fallibly so) that the aggregate well-being gain of those who benefit from *x* exceeds the aggregate well-being loss of those who are worse off. This argument assumes that we can make interpersonal comparisons of well-being: If Jim is better off in *x* and Sheila in *y*, then the change in Jim's well-being can be compared with the change in Sheila's. The argument also presupposes that money has roughly the same "utility" for Jim as for Sheila—that money in Jim's hands is transformed into well-being at roughly the same "rate" as money in Sheila's hands. If Jim is much richer than Sheila, then it is quite possible that some policy produces a small welfare benefit for Jim (relative to some baseline) and a larger welfare loss for Sheila, and yet Jim's WTP for the policy relative to baseline is greater than Sheila's WTA.

CBA's reliability as a proxy for overall well-being can be improved by "distributional" weights. Each person's WTP/WTA is multiplied by a weighting factor, inversely proportional to her wealth. Some academic work, and even governmental practice, takes this approach.

Preferences and Well-Being

CBA understands well-being as preference satisfaction. Such a view is standard in economics but still quite problematic. Someone's preferences might be hasty and unreflective. She might be insensitive to relevant facts. More subtly, her preferences might incorporate moral considerations. For example, if Sue on balance prefers *x* to *y*, even though her income is lower, because she judges the distribution of income in *x* to be fairer, Sue herself might quite possibly be worse off in *x*.

A more plausible view analyzes well-being in terms of idealized (fully informed, fully rational) and self-interested preferences and refines CBA by aggregating WTP/WTA amounts relative to such preferences. To be sure, specifying "full information," "full rationality," and "self-interest" poses major challenges.

Welfarism, Consequentialism, and Inequality

CBA (whether understood as an indicator of Kaldor-Hicks efficiency or, instead, of overall welfare) is consequentialist and welfarist in ranking outcomes. Welfarism says that if each person is equally well-off in two outcomes, the two are equally good. CBA satisfies this constraint (with well-being defined as preference satisfaction or as idealized and self-interested preferences with a matching refinement of the CBA test).

Consequentialism, at the level of governmental choice, says that government should evaluate policies in light of the goodness of their outcomes. CBA takes this route, and thus it is insensitive to deontological side constraints that arguably constrain the maximization of good consequences—side constraints barring torture, killing, deception, and so on.

Finally, CBA (even as corrected to incorporate distributive weights counterbalancing the variable "utility" of money) is an indicator of overall (total) well-being and does not take account of whether the distribution of well-being is more or less equal.

Challenges to CBA based on these concerns raise many complex issues. One response is to see overall welfare as *one* normative factor relevant to government choice, along with others: non-welfare goods, deontological side constraints, and fair distribution. CBA, on this view, is not a "superprocedure" but instead a rough proxy for *one* of the items (overall welfare) that government should care about.

Incommensurability

Many find it problematic that CBA employs money as a universal metric. CBA supposes that *whatever* the qualitative differences between two outcomes bearing on some individual's well-being, there is a single WTP/WTA amount for her that perfectly equilibrates those differences. Is this really true? For example, suppose that Fred has more friends in one outcome. Is the welfare value of these additional friends precisely equal to the value of any specific change in Fred's monetary assets? Isn't friendship valuable just because we resist thinking about its value in monetary terms? Suppose, tragically, that Fred loses a child. Can any sum of money compensate him for his loss?

The "incommensurability" critique of CBA subsumes these various challenges: (a) *incomparability* (that someone might be incomparably well-off in two outcomes rather than better off in one or equally well-off in both), (b) *constitutive incommensurability* (that the welfare value of certain goods consists in refusing to consciously monetize them), and (c) *lexical orderings* (that some losses are too great to be compensated by dollars).

CBA can be refined to deal with the incomparability challenge (although at the cost of additional complexity). CBA does require individuals to consciously monetize goods—but only occasionally, when posed a monetization question as part of a "stated preference" survey, designed to estimate WTP/WTA amounts. Nothing in CBA requires that individuals engage in ongoing monetary thinking throughout day-to-day life; and to insist that some goods are damaged if *ever* consciously monetized would seem to understate the resilience of these goods. Finally, CBA handles lexical orderings by shifting from the valuation of a good to the valuation of the risk of gaining or losing it. This is the approach used to arrive at WTP/WTA values for policies that save lives or cause deaths.

Matthew D. Adler

See also Homo Economicus; Pareto Optimality; Policy Applications of the Social Sciences; Rational Expectations; Social Choice Theory; Welfare Economics

Further Readings

Adler, M. D. (2012). *Well-being and fair distribution* (chaps. 2 & 3). New York, NY: Oxford University Press.

Adler, M. D., & Posner, E. A. (2006). *New foundations of cost-benefit analysis*. Cambridge, MA: Harvard University Press.

Mishan, E. J., & Quah, E. (2007). *Cost benefit analysis* (5th ed.). London, England: Routledge.

Covering-Law Model

The covering-law model is an early and influential model of scientific explanation that suggests that to explain a phenomenon is to show how a statement of that phenomenon can be derived using an argument that includes laws. Its chief advocate was the philosopher Carl Hempel. While the model was originally developed in the context of the natural sciences, Hempel and others attempted to extend it to the social sciences. A number of varieties of covering-law models have been developed, the most important of which are the *deductive-nomological* (DN) model and the *inductive-statistical* (IS) model. The DN model features deductive arguments and deterministic laws, while the IS model features inductive arguments and statistical laws.

As a simple example of a covering-law explanation, consider this explanation of why a helium-filled balloon will rise:

> The weight of the helium balloon is less than the weight of the volume of air the balloon displaces.
>
> Whenever an object weighs less than the air it displaces, *it will rise when released*.
>
> This balloon rises when released.

The explanation is in the form of an argument. The premises of an explanatory argument are called the *explanans,* while the conclusion is called the *explanandum*. What makes this argument a covering-law explanation is that the explanans contains a *general law*—here the claim that whenever an object weighs less than the air that it displaces, it will rise when released.

The following section presents a review of the requirements for and difficulties with the DN and IS variants of the covering-law model. The entry concludes with a discussion of some special difficulties with the application of these models to the social sciences.

The Deductive-Nomological Model

The DN model is the earliest and most widely discussed version of the covering-law model. According to this model, a DN explanation is an argument that meets the following four conditions:

1. The argument must be deductively valid.
2. The explanans must contain at least one law.
3. The explanans must have empirical content.
4. The statements comprising the explanans must all be true.

To say that the explanatory argument must be deductively valid is just to say that the explanandum is the logical consequence of the explanans—in other words, that it is impossible for all the premises in the explanans to be true and the conclusion to be false.

To clarify the requirement that the explanans must contain one law, one must be able to say just what counts as a law, and this has been a major challenge. As a starting point, philosophers of science

take laws to be certain kinds of universal generalizations, for example, "All metals conduct electricity." However, not all universal generalizations are laws. Laws must be distinguished from generalizations that are only accidentally true—for instance, the generalization that all the students in the philosophy class are less than 25 years old. One typical way of distinguishing laws from accidental generalizations is to say that the former, but not the latter, support counterfactual conditionals. Advocates of the DN model typically accept a Humean analysis of laws, understanding laws to be generalizations of unrestricted scope, whose terms do not refer explicitly or implicitly to particular objects, places, or times. This criterion rules out many obviously accidental generalizations, like the one about the age of students in the philosophy class, but it also rules out generalizations that one might consider laws. For instance, the generalization in the explanation of the balloon's rising is non-accidental, but it makes reference to the weight of an object, and thus implicitly to particular objects, namely, the earth and its gravitational field. The Logical Empiricists hoped that generalizations such as this might ultimately be derived from purely general laws, but this program of reduction has proved intractable. Moreover, there remains much debate about how many laws there are, if indeed there are any at all, and about what is an appropriate criterion for lawfulness.

The requirement that the explanans have empirical content is meant to rule out pseudoscientific explanations, in which the supposed laws or explanatory facts are not empirically testable. For instance, a Freudian explanation of a neurosis derived from laws concerning the interaction of the id and the superego would only meet this standard if it were possible to submit these supposed laws to at least an indirect empirical test.

Hempel's fourth condition, that the statements in the explanans must be true, may seem obvious, but it poses certain problems. For instance, because Newtonian mechanics have been superseded by relativistic mechanics, putative DN explanations of, for example, the motion of balls rolling down inclined planes will, according to this criterion, not be genuine. More generally, idealization and approximation are increasingly recognized as central to the explanatory practices of both the natural and the social sciences—and this is not apparently permitted within the DN model.

One consequence of the DN model (and covering-law models generally) is the structural identity of explanation and prediction. Hempel suggested that explanations and arguments share the same structure, with the difference only depending on contextual factors, such as whether the state of affairs described in the conclusion of the predictive/explanatory argument has yet occurred. For instance, the argument at the beginning of this entry can be viewed both as an explanation of why the balloon will rise and as a prediction that it will rise. However, critics of the DN model have argued that there are many instances where explanations do not predict and predictions do not explain. It is possible, for instance, to construct a DN argument inferring the height of a flagpole from the length of the shadow it casts and the angle of the sun. This argument can predict the height of the flagpole, but clearly it does not explain it. Conversely, it seems possible to explain events (especially low-probability events) after the fact even if it is not possible to predict them. A probable diagnosis of the difficulties with the structural identity thesis is that explanations must appeal to causes where predictions do not. The height of the flagpole is causally relevant to the length of the shadow, so it both predicts and explains that length, but since the length of the shadow is not causally relevant to the height of the flagpole, it predicts but does not explain it.

Another well-known counterexample to the DN model is this: Whenever a magician hexes a bit of salt and throws it into water, the salt dissolves. If the DN model is correct, an argument using this "law" will explain the salt's dissolving by reference to its having been hexed. As in the case of the flagpole and the shadow, what this counterexample seems to show is that the DN model has failed to capture the explanatorily essential concept of causal relevance.

The Inductive-Statistical Model

The IS model is Hempel's attempt to extend the covering-law model to the realm of statistical generalizations (or laws). Consider the following example:

> A patient with breast cancer who receives chemotherapy with drug A has a 95% two-year survival rate.
>
> *Smith received chemotherapy with drug A.*
>
> Probably, Smith survives two years.

This argument exemplifies the IS model of explanation, and the statistical generalization both predicts and explains Smith's survival. Formally, the requirements for IS explanations are identical to those for DN explanations, except that the laws involved are statistical and, accordingly, the arguments are inductive. Instead of requiring that the argument be deductively valid, the IS model requires that the argument be inductively strong—in other words, that it is improbable that the premises will be true while the conclusion false.

The IS model suffers from difficulties similar to those encountered in the DN model but with some added complications due to the addition of statistical premises. The argument above looks initially explanatory. But suppose it is the case that the two-year survival rate for breast cancer victims not taking the drug is also 95%; then, like the hexing of the salt, the taking of the drug is causally and explanatorily irrelevant.

An additional problem with statistical generalizations is that the probabilities involved often represent ignorance of relevant facts, so that with the addition of further facts, seemingly legitimate explanations turn out to be not explanatory. Suppose, in the above example, it is discovered that Smith's cancer is of a rare type that has a very low survival rate when treated with drug A. In that case, the argument would not explain Smith's survival.

A final problem has to do with Hempel's requirement that an IS explanation must show that the explanandum is probable in light of the explanans. Suppose, for instance, that Smith suffered from a virulent cancer strain with almost certain mortality. An experimental treatment promises a 5% chance of two-year survival, whereas in the absence of the treatment, the survival rate is near 0%. Even if Smith took the treatment she would be unlikely to survive. But even so, if she did survive, one would think that the treatment explained it. This example shows that in the IS model, as in the DN model, there turns out to be a breakdown of the structural identity between explanation and prediction.

Applications of the Covering-Law Model to the Social Sciences

The attempt to apply covering-law models to the social sciences is part of a larger project of metaphysically and methodologically unifying the natural and the social sciences. This project—often called *naturalism*—assumes that, at least ideally, social-scientific theories, like the theories of natural science, should consist of laws and that these theories can explain by showing how the laws cover the phenomena described by social science. Accordingly, much of the debate has been concerned with whether social-scientific phenomena are law governed or not.

It is sometimes held that there are no laws in the social sciences because the phenomena studied are too complex. Such laws as one does find in fields such as economics are ceteris paribus (all things being equal) laws—laws that hold only under idealized conditions, which are never realized and can seldom, if ever, be used to derive concrete predictions. One possible conclusion to draw from the ceteris paribus character of social-scientific laws is that naturalism is false, but increasingly, philosophers of natural science are inclined to think that laws in natural science are also ceteris paribus laws. Nancy Cartwright, in particular, has emphasized the parallels between explanations in physics and economics, arguing that in both cases there is a trade-off between truth and explanatory power. If Cartwright is correct, these cases reveal a defect in the covering-law model.

A second obstacle to the application of covering-law explanations is the supposed interpretive or hermeneutic character of explanation in the social sciences, whereby human actions in the social world are "understood," not explained, by subsumption under covering laws. Advocates of the interpretive approach suggest that it is impossible to explain human behavior and social institutions in purely empirical terms or to characterize the reasons that explain human action in terms of law-governed causes.

The difficulties in applying the covering-law model to the social sciences have led philosophers to seek other models more appropriate to actual explanatory practices. Two important alternatives are functional and mechanistic explanations. Functional explanations explain the existence of a social phenomenon by reference to its causal role or consequences. Mechanistic explanations seek to explain a phenomenon by describing the causal mechanism that produces it. Both of these forms of explanation seem to be important to social scientists, and neither appears to be reducible to covering-law explanation.

Stuart Glennan

See also Causation in the Social Sciences; Explanation, Theories of; Explanation Versus Understanding;

Hypothetico-Deductivism; Idealization in Social-Scientific Theories; Laws of Nature; Logical Positivism/Logical Empiricism; Mechanism and Mechanismic Explanation; Naturalism in Social Science; Reduction and the Unity of Science; Reductionism in the Social Sciences; Scientific Method

Further Readings

Cartwright, N. (1983). *How the laws of physics lie.* Oxford, England: Clarendon Press.

Hempel, C. G. (1965). Aspects of scientific explanation. In C. G. Hempel (Ed.), *Aspects of scientific explanation and other essays in the philosophy of science* (pp. 331–496). New York, NY: Free Press.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*(1948), 135–175.

Kincaid, H. (1990). Defending laws in the social sciences. *Philosophy of Social Science, 20,* 56–83.

Salmon, W. (2006). *Four decades of scientific explanation.* Pittsburgh, PA: University of Pittsburgh Press.

Criminology, Epistemological Critique of

This entry brings to light a cluster of epistemological and methodological issues regarding the scientific status and practice of criminology. These issues are related to methodological concerns in the philosophy of the social sciences as well as to issues in the cognitive sciences.

Criminology is the scientific study of crime and criminal justice systems. The field exists because some actions have been deemed "crimes," which merit a public response such as incarceration or rehabilitation. Because criminology adopts a scientific perspective on crime, it is distinct from theories of justice, which may examine what acts *should* be considered criminal. The basic schools of thought include the classical school, the positivists, the Chicago school, and the (newly evolving) modern schools of criminology.

Origins

The first text in criminology is generally considered to be Cesare Beccaria's 1764 work *On Crimes and Punishments* (*Dei delitti e delle pene*), which opposed torture and proposed that punishment be proportionate to the crime. Lawrence W. Sherman has noted, however, that Henry Fielding published an essay in 1751 calling for the rationalization of criminal justice, a state-run police force, and professional prosecutors. The "classical" school of Beccaria, along with the utilitarian philosopher and legal reformist Jeremy Bentham in the late 18th century, as well as others, saw crime as a rational choice responsive to incentives. The individual makes a simple calculation of whether the expected pain is greater than the expected pleasure. The solution from a society's standpoint is to create deterrence by providing swift and sure punishment.

The positivist school of criminology was founded in 1876 by Cesare Lombroso, who spoke of the "born criminal," whose "atavistic" (biologically "primitive") character was revealed by physical traits such as a sloping brow. The school emphasized empiricism and the scientific method to discover those aspects outside an individual's control that made the individual indulge in criminal behavior. Various researchers focused on the biological, social, demographic, or psychological causes of criminal behavior.

In the 1920s, theorists of the Chicago school shifted attention from the supposed biological or psychological causes of crime, adopting a "sociological ecology" approach that emphasized "social disorganization" as a cause of crime. This school sought to solve the problem of criminal behavior by changing the social environment in which the individual functioned.

Modern Developments and Epistemological Issues

Modern criminologists have begun to reject these more simplistic theories and have sought to create a more complex, cross-disciplinary view of crime in the context of modern society. Criminologists have adopted empirical methods, but there is some disagreement over what particular methods are the most appropriate to the subject. In particular, Lawrence Sherman and others extol an "experimental" approach as a necessary complement to the more "analytical" methods that may dominate the field today. Some writers take a critical stance and suggest that criminological theory and methods may be a function of class interest, political power, or similar considerations that are generally viewed as illegitimate in standard views of science.

An epistemological critique of criminology may be directed to criminal justice studies, asking questions such as "How do we know that our theories are right?" It may also be directed at criminal justice systems, asking questions such as "How do we know whether the suspect committed the crime?" At both these levels, the methodological and the scientific, relatively little has been written about the epistemology of criminology. At the methodological level, however, epistemological critiques of criminology can draw on the vast existing literature on the methodology of the social sciences. The epistemic properties of criminal justice systems have been less thoroughly explored in spite of evidence of error in both common law and civil law countries. The "reliabilist" or "veritistic" social epistemology described in Alvin Goldman's *Knowledge in a Social World* emphasizes truth conductivity and gives us a good foundation for the epistemological critique of criminology.

Cognitive biases cause persistent, recurrent, and systematic errors in criminal justice systems. Both criminal justice personnel and nonprofessional participants in the system are subject to cognitive biases that may lead to errors such as false convictions. Police officials are subject to confirmation bias, tunnel vision, and *us versus them* thinking. Forensic scientists are subject to cognitive biases that may encourage errors favoring the police theory of a case, as are prosecutors and judges. Suspects themselves may contribute to false convictions through false confessions, which are more likely in more vulnerable populations.

It seems clear that jurors are no less subject to cognitive bias than others. We do not know, however, whether deliberation improves outcomes. The infirmities of eyewitness testimony are probably better understood. Police lineups, for example, invite error if not properly structured. Eyewitnesses are more likely to misidentify a person perceived to be of a different "race."

Individual cognitive bias is not the only cause of error in criminal justice systems. The organization of a criminal justice system may influence the chance of error even when individuals are "rational" in some strong sense, such as that of economic theory. Police and prosecutors may have an incentive to maximize convictions rather than to separate the guilty from the innocent. Such incentives run afoul of the "multitask problem" of organizational economics, whereby unobserved aspects of the job are sacrificed to increase measured values of the observed dimensions. Police agencies may sometimes act to increase case clearings, without adequate regard for crime reduction or avoidance of false conviction.

The police interest in clearing cases may spread to the crime lab if it is organized under the police, as is typical in the United States. Lab fundings may depend in part on case clearing. In some cases, a lab may receive a payment for each guilty plea or guilty verdict involving lab testing or, perhaps, in certain categories such as drug use or drunk driving. Individual forensic scientists have sometimes been led into error in part because of a feeling of identification with the police.

The risk of forensic science error is increased by a twofold monopoly in forensic science. First, evidence is typically examined by one crime lab only. In this sense, the crime lab receiving a bit of evidence has a monopoly on examination of that evidence. Second, that same lab will normally be the only one to offer an interpretation of the results of the examination it performs. The monopoly structure of forensic science contrasts with the relatively open competition in research science, where any one lab may challenge the results of any other lab.

The monopoly structure of forensic science interacts with the architecture of the human mind. Even if the individual scientist's objective is to provide unbiased information, he or she is still an agent of the police and the prosecution. Moreover, standard practice gives the examiner context information about the crime, the victim, and the suspect. These conditions create conscious or unconscious bias that may induce error.

Different lines of evidence, such as eyewitness testimony, confessions, and different types of forensic evidence, will often emerge from one investigation and one team of investigators. This unity of the evidence-generation process creates the risk that errors will be correlated across evidence classes. Police investigators may induce a false confession from a suspect, employ a snitch to testify against the suspect, consciously or unconsciously influence eyewitnesses to testify against the suspect, and provide the crime lab with context information that may induce an incriminating forensic science error. In such cases, errors are correlated across seemingly independent evidence channels. Jurors, who implicitly assume

that errors are independent across types of evidence, may underestimate the probability of innocence.

If the epistemic infirmities of criminal justice systems are greater than is generally recognized by affected citizens, there may be cause to consider reform measures such as taping suspect interviews, enhancing the defense's right to expertise, separating crime labs from law enforcement, and eliminating the snitch system. The relatively unambiguous distinction between guilt and innocence may also help make criminal justice systems a good area for further research in social epistemology.

Roger Koppl and Evard James Cowan

See also Cognitive Sciences; Epistemology; Law, Social Phenomenon of; Legal Epistemology; Philosophy of Expertise; Social Epistemology

Further Readings

Goldman, A. (1999). *Knowledge in a social world.* Oxford, England: Oxford University Press.

Koppl, R. (2005). How to improve forensic science. *European Journal of Law and Economics, 20,* 255–286.

Laudan, L. (2006). *Truth, error, and criminal law: An essay in legal epistemology.* New York, NY: Cambridge University Press.

Martin, D. (2002). Lessons about justice from the laboratory of wrongful convictions: Tunnel vision, the construction of guilt, and informer evidence. *UMKC Law Review, 70,* 847–864.

Miller, M. J. (2009). *21st century criminology: A reference handbook*. Thousand Oaks, CA: Sage.

Sherman, L. W. (2005). The use and usefulness of criminology, 1751–2005: Enlightened justice and its failures. *Annals of the American Academy of Political and Social Science, 600,* 115–135.

Critical Rationalism

Critical rationalism is a term sometimes used to refer to Karl Popper's philosophical system, or a large part thereof. But it more precisely refers to a critical *attitude* that Popper advocated, and derivatively to the claim that we ought to adopt this attitude (in a significant class of contexts, e.g., when doing science).

In *The Open Society and Its Enemies,* Popper contrasts critical rationalism with comprehensive rationalism. Each position involves the view that we should employ argument and experience in order to assess statements; this is the *rationalist* component. (Do not be confused by the fact that "rationalism" is typically used in a different sense in contemporary philosophy, to refer to the view that pure reason, unaided by experience, is a source of knowledge—standardly taken as the opposite of "empiricism.")

This entry presents the position of critical rationalism by distinguishing it from other kindred rationalist stances, presents critiques of it, and ends by showing the relevance as well as the limitations of rationalist stances, both in the case of an individual versus a group she or he is a member of and as regards how scientists should be expected to proceed rationally, or not, in benefiting science.

Comprehensive Rationalism

The comprehensive rationalist follows (or strives to follow), implicitly or explicitly, the rule that *any statement that cannot be supported by argument or experience should be abandoned.* And there is scope to understand "abandoned" in several different fashions, such as "classified as false," "believed to be false," "not classified as true," and "not believed to be true." (Failing to believe that a given statement is true is not equivalent to believing that it is false; one might suspend judgment as to its truth or be unaware of the statement and thus be unable to form an opinion on its veracity.) So the classical epistemological position that one should not believe in something without support from reason or experience is a form of comprehensive rationalism.

The problem with comprehensive rationalism, however, is that it appears to be self-refuting. Consider the aforementioned statement: "Any statement that cannot be supported by argument or experience should be abandoned." As *this* statement cannot be supported by argument or experience—this is typically uncontested in the literature—it should be abandoned even if it is true! And while it would be easy to achieve a quick fix by changing "any statement" to "any statement except this one," it would be ad hoc. To foreshadow the subsequent discussion, the comprehensive rationalist

would have difficulty in explaining why one should not instead adopt a foundational rule, such as "Any statement that cannot be supported by appeal to the King James Bible—except this one—should be abandoned."

Critical Rationalism

In light of the problem with comprehensive rationalism, Popper suggests that the rationalist attitude must ultimately rest on faith (or irrational commitment). And he suggests that it is therefore appropriate to adopt a more modest stance, a *critical rationalism,* where this is frankly admitted. As part of this, Popper advocates a critical attitude of accepting that one might be wrong and that those with whom one disagrees may be right. He also suggests that we should work together to eliminate errors.

However, critical rationalism, and more particularly the view that it is acceptable for rationalism to be based on faith, is criticized in turn by William Warren Bartley. The problem is that if the rationalist embraces faith as her starting point, then she concedes a priority to faith. So anyone who does not want to be a rationalist is excused, and there is no loss of "intellectual integrity" in opting for a form of irrationalism instead. Imagine, for example, that a critical rationalist were to criticize a religious person for summarily dismissing the results of a well-developed special science, such as physics. This person could simply respond with "But you, too, start with faith. And my alternative faith leads me to believe that these results are false." In fact, Bartley's *The Retreat to Commitment* pays particular attention to how this tu quoque move could be employed by Christians.

Pancritical Rationalism

Bartley therefore suggests that critical rationalism may be replaced with *pancritical* rationalism (which is also known as *comprehensively critical rationalism*). The simple idea is that the rationalist may forswear any reliance on faith whatsoever by holding all her positions open to criticism and by being willing to give any of them up under the right circumstances. Crucially, this goes for pancritical rationalism *itself;* in essence, the idea is that argument and experience might be used, in principle, to show that we require faith instead.

A radical component of Bartley's view is that it can be perfectly rational to believe in something—or to hold some position—even in the absence of any justification (in the sense of a reason) for doing so. On the contrary, the rationality in holding a position consists precisely in the willingness to subject it to criticism and to abandon it if it fails to stand up under fire. This fits nicely with the emphasis on testing in Popper's philosophy of science and with the idea that it's what we do with our theories, when we have them, that really matters.

It should be noted that simply avoiding susceptibility to the tu quoque move does not provide an argument *for* being a pancritical rationalist. Such an argument is supplementary and might be made on either ethical or epistemic grounds; in particular, it may be urged that being a pancritical rationalist (or as close to the ideal as one can be) is virtuous. Nonetheless, there is a gap between what is best for an individual and best for the group of which she is a part. And there is reason to think that science benefits from having some dogmatic scientists and does rather better in virtue of their presence than it would if *all* scientists were pancritical rationalists.

Darrell P. Rowbottom

See also Context of Discovery Versus Context of Justification; Epistemology; Popper's Philosophy of Science; Scientific Method; Social Epistemology; Virtue Epistemology

Further Readings

Bartley, W. W. (1962). *The retreat to commitment.* New York, NY: Knopf.

Jarvie, I. (2001). *The republic of science: The emergence of Popper's social view of science, 1935–1945.* Amsterdam, Netherlands: Rodopi.

Miller, D. W. (1994). *Critical rationalism: A restatement and defence.* Chicago, IL: Open Court.

Popper, K. R. (1945). *The open society and its enemies.* London, England: Routledge.

Rowbottom, D. P. (2011a). Kuhn vs. Popper on criticism and dogmatism in science: A resolution at the group level. *Studies in History and Philosophy of Science, 42,* 117–124.

Rowbottom, D. P. (2011b). *Popper's critical rationalism: A philosophical investigation.* London, England: Routledge.

Critical Realism in Economics

The central claim of the project of critical realism in economics is that the discipline would benefit from a more explicit, systematic, and sustained concern with *ontology*—that is, with the philosophical analysis of the nature of what exists in the social world—than its practitioners have hitherto displayed. Underpinning this claim is the belief that social research is most likely to bear fruit if it uses tools that are tailored to suit the nature of the material under investigation. To that end, critical realists style themselves as adopting an underlabor role, whereby they develop an account of the ontology of the social world that is used to inform an analysis of the research methods that are most likely to bear fruit in social research.

Main Tenets

Holism, Structures, and Openness

Critical realists argue that many commonly acknowledged general features of the social world—such as the possibility of genuine human choice, the often routinized nature of (much of) social life, the segmented and other-directed nature of many social practices, and the fact that the practices followed in particular settings often remain the same even when the specific individual carrying them out changes—can be rendered intelligible only if it is acknowledged that the social world has certain broad features. These features show that the social world is structured, consisting not only of people but also of social structures—social rules and institutions—that are ontologically irreducible to social practices. The social world is therefore in large part holistic, with the aforementioned structures consisting of networks of internally related positions (e.g., those of capitalist and worker, buyer and seller, and creditor and debtor) through the occupancy of which people and social structures come together. Furthermore, the social world is shown to be inherently dynamic or processual—with people drawing on social structures in order to act and, in acting, either reproducing or transforming both those structures and their own capacities and personalities—and open, being characterized by a paucity of sharp, stable event regularities of the form "Whenever this event or state of affairs *x*, then that event or state of affairs *y*."

On this view, socioeconomic phenomena are to be explained as the outcome of the causal interplay over historical time between (antecedent) social structure and (subsequent) human agency. More specifically, the initial stage of an explanation involves the identification of the practices responsible for the phenomenon under investigation, after which it is necessary to uncover the social structures and tacit skills that facilitate those practices, together with any conscious and unconscious psychological factors that motivate them. Within this general view, the distinctive role of the discipline of economics is that of being the branch of social science that focuses on the social structures and processes governing the production, distribution, and use of the material conditions of well-being.

Causal Mechanisms

Epistemic access to these underlying causal mechanisms can often be obtained by focusing on what critical realists term "contrastives"—that is, situations where there are two or more comparable groups or populations with similar histories and shared conditions and where our background knowledge leads us to expect a specific relationship (often, though not always, one of similarity) between the outcomes experienced by those groups but where we are surprised ex post facto by the relation actually discovered. For instance, the contrast in question might stem from the observation that since World War II productivity growth has often been lower in the United Kingdom than in the economies of other major industrialized nations. The existence of such contrasts provides prima facie evidence that there is an unidentified causal mechanism at work, whose influence accounts for the unexpected relation between the outcomes displayed by the populations or groups in question. And, having been alerted to the (potential) existence of a hitherto unknown causal mechanism in this way, the task of the researcher is then to identify and illuminate it—perhaps by investigating the hypothesis that the ill-understood mechanism in question is analogous to some other, better-known mechanism in a different field of inquiry—thereby explaining the contrast in question.

Critical Realism Against Mainstream Economics

One key implication of their social ontology, critical realists contend, is that mainstream economics is unlikely to yield insights into the social world. For critical realists, the hallmark of mainstream economics lies in its commitment to formal mathematical modeling and to a mode of explanation according to which social phenomena of interest are explained when a description of them is deduced from sets of axioms and assumptions expressing regularities of the following form, as we have seen: "Whenever this event or state of affairs (x), then that event or state of affairs (y)." However, the a posteriori observation—in particular, of the fragility of many of the results obtained by econometricians—reveals that the socioeconomic world displays few sharp, stable event regularities of the kind required for such deductivist explanations. What this implies is that the construction of the orthodoxy's mathematical-deductivist models requires that economists make assumptions—for example, that people always and everywhere maximize their expected utility or that the economy consists only of one "representative" agent—designed to ensure that under given conditions (x) the homunculi who inhabit the model economic worlds in question always act in the same way and produce the same (probability distribution of) outcome(s) (y) as the deducibility of theoretical results requires. The problem with such assumptions is that they are not descriptive of most of the social world; designed as they are to eliminate various sources of indeterminacy from models of a social world that is in large part open, such assumptions portray people as passive atoms who inhabit isolated or self-contained worlds, when in fact they are nothing of the sort. The scope for such fictional models to advance our understanding of real economic issues is, Tony Lawson argues, far from obvious and certainly insufficiently well established to warrant the way in which their use dominates the economics profession.

Relation of Critical Realism With Heterodox Economics

On the other hand, there is a good deal of common ground between critical realism and the views held by various heterodox schools of economic thought. Many Austrian, feminist, old-institutionalist, Marxist, and post-Keynesian economists make the case for their own preferred style of economic analysis by using arguments that presuppose, if only implicitly, a social ontology closely akin to that espoused by critical realists. The value of critical realism for heterodox economists lies in the way it affords them philosophical resources that can help them make explicit, clarify, systematize, and so realize the full potential of their insights into the nature of the socioeconomic being and its implications for an appropriate methodology for economics, in particular by facilitating the clarification and refinement of various important substantive concepts, including *social order, probability, money, markets,* and *capabilities.*

Paul Lewis

See also Analytical Sociology and Social Mechanisms; Austrian Economics; Causation in the Social Sciences; Feminist Economics; Heterodox Economics; Holism, in the Social Sciences; Institutional Economics; Marxist Economics; Philosophy of Economics, History of; Realism and Anti-Realism in the Social Sciences; Social Ontology, Recent Theories of

Further Readings

Fullbrook, E. (Ed.). (2009). *Ontology and economics: Tony Lawson and his critics*. London, England: Routledge.

Lawson, T. (1997). *Economics and reality.* London, England: Routledge.

Lawson, T. (2003). *Reclaiming reality.* London, England: Routledge.

Lewis, P. A. (Ed.). (2004). *Transforming economics: Perspectives on the critical realist project*. London, England: Routledge.

Cultural Evolution

Culture involves the social transmission of novel behavior, such as gestures, songs, or tool-making skills, among individuals who are generally members of the same species. Elements of culture spread both vertically, from one generation to another, and horizontally, among members of a generation. Thus, two key components of culture are (1) a means of *generating* novel behavior and (2) a means of *spreading* it, such as imitation and other forms of social learning. This entry outlines efforts to develop a scientific

framework for understanding the dynamics of cultural change.

It is widely believed that nonhuman species such as chimpanzees possess culture, but human culture is noteworthy because it is *cumulative*. Individuals build on each other's ideas, not just in a random manner that reflects mere copying errors, individual differences, and limitations in physique or differences in the materials at hand, but in a way that is *adaptive*. We strategically modify the ideas of others to suit our own needs and tastes, such that cultural products become more useful, expressive, or aesthetically pleasing. Anthropologists usually place the origins of human culture approximately 2 million years ago, during the transition from *Homo habilis* to the upright walking *Homo erectus*. Our earliest stone tools date back to this time. However, there is little evidence of *cumulative* cultural change until approximately 100,000 to 50,000 years ago, during the Middle-Upper Paleolithic Era. This period constitutes a dramatic transition in human culture, as it marks the beginning of many kinds of cultural artifacts, including task-specific tools, art, ornamental jewelry, and indications of ritualized religion.

Thus, not only does adaptive cultural change accumulate over time, but it *diversifies,* becomes increasingly complex, and exhibits phenomena observed in biological evolution, such as niches, and punctuated equilibrium. Like biological evolution, culture is *open-ended*; there is no apparent limit to the variety of new forms it can give rise to. For these reasons, it has been argued that culture constitutes a second evolutionary process, one that, though it grew out of biological evolution, exhibits an evolutionary dynamic in its own right that cannot be reduced to biology. Some of what is considered cultural behavior *can* be explained by biology. Most, however, would probably concede that, much as principles of physics do not go far toward an explanation of, say, the vertebrate body plan (though things like gravity play some role), biology does not go far toward an explanation of, say, the form and content of a sonnet (though factors like selective pressure for intelligence play some role). Most would concede that to explain how and why such forms arise, accumulate, and adapt over time, one must look to culture. Nevertheless, the attempt to establish an evolutionary framework for culture remains a struggle.

One theory is that culture evolves through natural selection. What necessitated Charles Darwin's theory of how species evolve through natural selection is that traits that are *acquired* by an organism over its lifetime are not transmitted to the offspring and thus not passed down through time. Thus, whereas a rock that smashes stays smashed, if a rat loses its tail, the rat's offspring are *not* born tail-less; the "bitten-off tail" trait is lost from the rat lineage. The paradox Darwin faced was this: If change is not retained from parent to offspring, then how does change *accumulate?* His solution was as follows: He proposed that part of an organism *is* retained (the part we now call DNA). It is responsible for traits shared by relatives, and it undergoes random change from one generation to the next. Since random changes that are beneficial cause their bearers to have more offspring, or be selected, over generations, such traits proliferate at the expense of detrimental ones and become more widely represented in a population.

John von Neumann determined the minimal algorithmic structure capable of evolving through natural selection, which he termed a *self-replicating automaton*. It consists of self-assembly instructions that are both *copied* to make offspring and *interpreted* to develop them into adults. However, it was the simpler notion of a *replicator,* something that merely makes copies of itself, that inspired the *Darwinian theory of culture*. According to this theory, culture evolves through the variation and selection of replicators, sometimes referred to as *memes*. Darwinian anthropologists apply phylogenetic methods, developed for classifying biological organisms into lineages, to cultural artifacts. This works well for highly conserved assemblages but falsely classifies similarity due to horizontal exchange of ideas as similarity originating from a common ancestor. *Dual-inheritance theory* is a version of the Darwinian view that posits that humans have two inheritance systems—biological and cultural—and focuses on processes that bias the transmission of cultural information, such as the tendency to preferentially imitate high-prestige individuals. However, since the theory of natural selection assumes that variation is randomly generated and acquired change is not transmitted, to the extent that transmission is biased from the random, and ideas acquire change between transmission events as we contemplate them, the Darwinian approach gives a distorted model.

Another theory of cultural evolution was inspired by recent evidence that the earliest forms of life were

self-organized metabolic networks that evolved through a non-Darwinian process involving transmission of acquired traits, sometimes referred to as *communal exchange*. It has been proposed that what evolves through culture is *worldviews*—the integrated webs of ideas, beliefs, and so forth, that constitute our internal models of the world—and they evolve, as did early life, not through competition and survival of the *fittest* but through communal transformation of *all*. In other words, the assemblage of human worldviews changes over time not because some replicate at the expense of others, as in natural selection, but because of ongoing mutual interaction and modification. Elements of culture, such as rituals, customs, and artifacts, reflect the states of the worldviews that generate them. The theory is consistent with network-based approaches to modeling trade, artifact lineages, and the social exchange of knowledge and beliefs and with the unexpectedly high degree of cooperativity in human culture.

Liane Gabora

See also Biology and the Social Sciences; Cooperation, Cultural Evolution of; Evolutionary Psychology; Human Cultural Niche Construction and the Social Sciences; Naturalism in Social Science; Primatology and Social Science Applications; Sociobiology

Further Readings

Boyd, R., & Richerson, P. (2005). *The origin and evolution of cultures*. Oxford, England: Oxford University Press.

Gabora, L. (2008). The cultural evolution of socially situated cognition. *Cognitive Systems Research, 9*(1–2), 104–113.

Mesoudi, A., Whiten, A., & Laland, K. N. (2006). Towards a unified science of cultural evolution. *Behavioral & Brain Sciences, 29,* 329–383.

Cultural Studies

Cultural studies can be seen as an extension of and, at the same time, a challenge to the social sciences. This is because it has a highly ambiguous relation to them, with some practitioners of cultural studies seeing themselves as social scientists (or akin to them) but many being wary of the scientific claims made by them. This relation is also complicated by the broad and multidisciplinary (or sometimes antidisciplinary) character of cultural studies, which is highly dispersed geographically and not easy to define in terms of either its objects of study or its methodologies. Of course, this does not mean that it does not have dominant themes and common ways of thinking, but it does mean that it does not have a simple definition.

The most convenient way of understanding cultural studies and its theoretical evolution is to refer to one of its key founding institutions in terms of its later development and expansion—the Birmingham Centre for Contemporary Cultural Studies in England, which was founded by Richard Hoggart in 1964, a key early figure in the area, who distinguished himself in 1957 by publishing his *The Uses of Literacy.* Although Hoggart did not make any claims to being a social scientist, this publication has often been seen as an ethnographic approach to the description, analysis, and understanding of working-class life. This work, along with E. P. Thompson's 1963 work, *The Making of the English Working Class,* and a whole series of studies by one of the major early scholars associated with the field, Raymond Williams, helped establish one of the main thematic areas that would be explored in what would become known as cultural studies—the study and analysis of working-class life and the values, beliefs, language, tastes, institutions, leisure interests, and social and political struggles that characterize it (although some critics now question the value of using a term like *the working class*). Nevertheless, these writers, and especially Raymond Williams (who theorized culture "as a whole way of life" from a materialist point of view), have become founders of the area.

This tradition of analysis would help stimulate many studies centered on popular culture and the understanding of youth subcultures, which would sometimes involve borrowing ideas from sociology (particularly the Chicago school). It might be noted here that early interest in working-class culture in Britain was not restricted to the Birmingham Centre; many extramural courses were taught outside the mainstream university system, and there were other centers in other parts of Britain (and sociology departments in many parts of the world). Also, the early stimulus for interest in popular forms of working-class culture came about as a challenge to what was often considered elitist views about "high/minority" versus "low/mass/popular" culture,

promulgated in what has become known as the Culture and Civilization tradition, associated with 19th-century writers such as Matthew Arnold and his influence on 20th-century critics including F. R. and Queenie Leavis and Denys Thompson. It was Raymond Williams who helped promulgate the anthropological idea that culture, far from being the preserve of cultural elites, was "ordinary" and would include everything from Dostoyevsky, Picasso, and Beethoven to sports, social clubs, family life, and entertainment.

The directorship of the Birmingham Centre was taken over in 1969 by Stuart Hall, a key figure in the New Left in Britain, and this was the moment when the Centre began to consider its relation to theory in very self-conscious ways. One of the key currents of thought (which can be broadly defined as philosophical) that has structured much writing in cultural studies is the Marxist tradition. However, in Birmingham, and beyond, the Marxist legacy has been complex and conflicted. Thompson was a Marxist who stressed working-class agency through its radical history; however, this was only one way of applying Marxist-inflected thought. One early way in which Marxist thought was mobilized was through the Frankfurt school's fusion of Marxist and Freudian thinking (referred to as "critical theory") to show how the culture industry dehumanized, disempowered, and depoliticized the working class, but this way of thinking (mainly associated with Theodor Adorno and Max Horkheimer) was complemented and challenged by other Marxist approaches.

The work of the French Marxist philosopher Louis Althusser exercised considerable influence over cultural studies' practitioners in the 1970s and 1980s, helping scholars theorize ideology with relation to what he called "state apparatuses" but particularly with his notion of "interpellation," which described how people actually become subjects of ideology. This had a particular influence on film theory as practiced by many writers who contributed to the journal *Screen,* which explored how audiovisual forms situated (or interpellated) the public as viewing subjects. One of the most enduring essays written during this period was the feminist film critic Laura Mulvey's "Visual Pleasure and Narrative Cinema" (first published in 1973), which argued that classic Hollywood cinema was biased toward the male gaze insofar as most of the protagonists were male, and women actors were reduced to their role of objects of male desire. In this way, the audience was interpellated in a thoroughly sexist way. Althusser also introduced the term *problematic,* which helped critics explore not only what a text says but also what lurks in the gaps and silences of *what it does not say,* so as to be able to unearth and explore the hidden ideological background as a kind of Freudian symptom.

One of the most important theoretical developments in the Birmingham Centre (and beyond) was the use of Antonio Gramsci's notion of *hegemony,* which describes the exercise of power within democracies as not only a question of the use of force (in relation to the police, the army, the courts, etc.) but also in relation to the role of how groups gain power through alliances, negotiation, and compromise in order to gain intellectual, cultural, and political leadership. This approach takes the emphasis off the more simplistic Marxist notion of the dominant class exercising power over the dominated in a more or less determined way and helps critics consider how power blocs are *themselves* subject to internal conflict and alliances. Here, culture is understood as the site of struggle where dominant values and ideas are subject to counterhegemonic strategies. This was linked to Gramsci's idea of the "organic" intellectual, a member of a repressed but revolutionary class who would create and propagate counterhegemonic ideas and values to challenge those of the dominant political order (something Stuart Hall felt was part of the role of scholars at the Birmingham Centre). These ideas, together with the application of hegemony to describe how official high forms of culture are established, became so important that historians of cultural studies often refer to it as *the turn to Gramsci.* Here, it is possible to see that the use of Marxist theory provided paradigms for analysis and helped forge politically motivated scholars who not only analyzed the world but also wished to change it (an important inheritance passed on to many [but not all] practitioners of cultural studies).

Other important theoretical currents that were explored in Birmingham and many other centers from around the 1980s (when cultural studies' institutional expansion began in earnest not only in Britain but, principally, in the United States and Australia and then in many other parts of the world) were related to what are referred to as structuralism, poststructuralism, and postmodernism. The turn to structuralism in the humanities, which was inspired by the work of linguists like Ferdinand de

Saussure, Charles Peirce, Roman Jakobson, and Émile Benveniste and the anthropologist Claude Lévi-Strauss, led to many attempts at what were considered "scientific" analyses of culture that put the emphasis on the way cultures could be explained with relation to the way complex social rules provided a "grammar" that explained all kinds of social phenomena with relation to codes and signs (Saussure provided a key theory of the sign, which would have enduring influence).

These ideas were developed by writers such as Roland Barthes and Umberto Eco and, in the case of Barthes's early work, were often linked to efforts to show how capitalist cultures were dominated by bourgeois values and interests (Althusser would draw on structuralist ideas as well as on Freud and Marx). After the first wave of structuralism, writers became more skeptical of the so-called objectivity of structuralist science and many turned to the later Barthes and the deconstructive philosophy of Jacques Derrida. Derrida's philosophy, known as *deconstruction,* undermined structuralism (and just about all forms of thought with claims to scientific truth) by careful readings that showed how all texts were ultimately structured on oppositions and rhetorical devices that were vulnerable to the destabilizing tactics of close reading. In this way meaning, while possible, is constantly differed and plural. Like many areas in the humanities, cultural studies began to draw on these ideas not in order to forward deconstructive philosophy *as such* but in order to question the discourses that had structured Western thinking. This interest in poststructuralist reading (as it became known) coincided with other theoretical and thematic interests that extended an interest in class to include subordination on the grounds of gender, race/ethnicity, and, later, sexuality.

This very broad interest in subordinated groups led to complex readings that also combined with Marxism, psychoanalysis, and theories and approaches developed within feminism, postcolonial theory, and queer theory. The ideas of the poststructuralist psychoanalyst Jacques Lacan have also been enormously influential in terms of the construction of identity. The Lacanian tradition does not assume that identity is the unique essence of the self found in much thinking in the Western philosophical tradition, including Christian notions of the soul, but instead it depends on symbolic cultural constructions. This means that the self does not precede culture and language as an essence but is actually a product of them. This idea, very reminiscent of structuralist approaches, has been applied to all kinds of identities whether they be personal, cultural, national, or historical.

Another major influence has been the work of Michel Foucault and his theories and methods concerning the construction of the lunatic, the criminal, and the homosexual in the human sciences. His insistence on the importance of "discourse" to the understanding of identity has often provided an alternative to the Marxist-Freudian traditions of thought. The work of Derrida, Lacan, and Foucault has often been associated with the idea of postmodernism, which generally questions, or complicates, the fixed meanings associated with philosophical and scientific traditions that seek absolute, stable truths and meanings. Two key thinkers in this context are Jean-François Lyotard and Jean Baudrillard. One of Lyotard's central claims was that from around the 1970s all the old certainties in thought were no longer credible and these were being replaced by "little narratives" that made no universalizing claims to truth and that truth was being replaced by *performativity* (a pragmatic approach to knowledge that only insists on utility rather than on notions of truth). This was part of what Lyotard referred to as "the postmodern condition." One of the most influential claims that Baudrillard made was that societies dominated by contemporary capitalism had lost contact with a sense of the real—they were "hyperreal," meaning that signs were no longer taken as referring to any kind of reality but were empty copies of copies.

Since the later part of the 20th century, there have been a number of shifts in the theoretical emphasis in cultural studies, with major contributions coming from critics who have placed the emphasis on empirically based audience studies that, in a loose philosophical sense, are related to phenomenology and reception theory. Furthermore, prominent critics like Stuart Hall and Lawrence Grossberg have suggested that cultural critics put less emphasis on identity politics and on questions of ideology, which tended to dominate cultural studies in the 1980s and 1990s. This has led to some innovative, if tentative, explorations of the work of philosophers like Gilles Deleuze and Félix Guattari, Giorgio Agamben, Slavoj Žižek, and Alain Badiou, and these explorations have been complemented by a thematic broadening that has seen studies focused on new media, anticapitalism,

the transnational, celebrity cultures, and digital and convergence cultures (among many others).

The field of cultural studies, then, continues to remain enormously eclectic from both the theoretical and the methodological points of view, challenging mainstream social sciences, erasing disciplinary boundaries, and drawing on all the writers mentioned above and many others, on a very broad-ranging and ever-expanding area.

David Walton

See also Frankfurt School and Critical Social Theory; Identity, Social; Ideology; Interdisciplinarity; Multiculturalism; Postcolonial Studies; Postmodernism; Structuralism and Poststructuralism

Further Readings

Barthes, R. (1972). *Mythologies*. New York, NY: Noonday Press. (Original work published 1957)

Barthes, R. (1985). *The fashion system*. London, England: Jonathan Cape. (Original work published 1967)

Grossberg, L., Nelson, C., & Treichler, P. (Eds.). (1992). *Cultural studies*. London, England: Routledge.

Hall, G., & Birchall, C. (2006). *New cultural studies: Adventures in theory*. Edinburgh, Scotland: Edinburgh University Press.

Hall, S. (1996). Cultural studies and its theoretical legacies. In D. Morley & K.-H. Chen (Eds.), *Stuart Hall: Critical dialogues in cultural studies* (pp. 262–275). London, England: Routledge.

Hoggart, R. (1957). *The uses of literacy*. Harmondsworth, England: Penguin.

Storey, J. (2009). *Cultural theory and popular culture: An introduction* (5th ed.). Essex, England: Pearson Prentice Hall.

Thompson, E. P. (1968). *The making of the English working class*. London, England: Penguin.

Thompson, E. P. (1995). *The poverty of theory*. London, England: Merlin. (Original work published 1978)

Turner, G. (2003). *British cultural studies: An introduction* (3rd ed.). London, England: Routledge.

Walton, D. (2012). *Doing cultural theory*. London, England: Sage.

Williams, R. (1980). *Problems in materialism and culture*. London, England: Verso.

Williams, R. (1992). *The long revolution*. London, England: Hogarth. (Original work published 1961)

Death and Immortality, Philosophical Perspectives

This entry provides a brief account of how philosophy looks at death by sketching some views about the nature of life, persons, death, and the implications for immortality.

Being *immortal* is not the same thing as being indestructible. Something is immortal if it *will* merely remain alive forever, which is not to say that it *could* not perish. What death is and whether immortality is possible, whether for us or for other sorts of beings, depends on what we are and what it is to be alive.

Life

To define "alive," we want to identify some capacity that is distinctive of living things. However, there are many candidates. On Aristotle's list, we find nutrition, appetite, growth, reproduction, perception, motion, and thought; later, biologists added responsiveness to the environment, homeostasis, and metabolism. It is not easy to define "alive" because it is difficult to find one characteristic that all and only living things exemplify. Aristotle's solution was to equate being alive with possessing *one or more* of the capacities on his list; unlike him, however, we know how to build mechanisms that possess some of these features yet are not alive.

Although no definition of life has achieved consensus, the only living things we know of are organisms and perhaps some of their components (e.g., organs), and there is wide agreement that an organism is alive only if it has the capacity to develop or maintain itself in a suitable environment. The processes by which it does this are called *vital processes*.

Persons

But it is not clear that you and I *are* organisms. Suppose that, as René Descartes said, we are essentially thinking things. The existence of thinking things need not depend on vital processes. If it is possible to build mechanisms capable of thought, such as the fictional android Data, then thinking beings need not be organisms and their persistence may not require that they be alive.

Some contemporary theorists, such as Peter Van Inwagen, Paul Snowdon, and Eric Olson, say we are organisms, but many others disagree. According to Derek Parfit and Jeff McMahan, we are minds, with psychological persistence conditions. Parfit has argued that we remain in existence over time if our psychological attributes do not suddenly change drastically and we do not undergo fission like an amoeba. Given his view, we need not be organisms to survive; in theory, we could transfer our psychological features to machines and continue our existence indefinitely. But this gives us the potential for immortality only if the notion of immortality is extended to embrace eternal existence (as opposed to eternal life).

Death

The term *death* is ambiguous. Death might be a *state*—the state we are in once we perish. Alternatively, it might be the *process* of extinction.

It could also be equated with one of two *events*: the point when the dying process reaches irreversibility or when it completes itself. The process of death can be compared to a race, which begins at some time, continues a while, and then ends; so it is not instantaneous (it takes time to complete itself).

It seems clear that only organisms and some of their components can die and also that an organism dies only when its vital processes cease. But many questions remain: (1) How is dying related to ceasing to exist? (2) Vital processes can be suspended, or even reversed; is an organism alive whose vital processes are suspended? (3) Is it possible to resurrect an organism that has died? (4) Is there some other way to avoid ceasing to exist?

Question 1 arises since it is conceivable that thinking beings (such as Data, from the TV series *Star Trek: The Next Generation*) could exist that are not alive at all. Such beings cannot die, but they could cease to exist; their ending would be deathless. Some theorists would add that dying does not entail ceasing to exist. Fred Feldman argues that we are organisms, but organisms do not cease to exist when they die. They remain in existence for a time—as corpses. On this view, we could continue our existence indefinitely, if not forever, as frozen corpses. However, many other theorists reject this position on the grounds that an organism ceases to exist when it dies.

Seeds, spores, some small creatures, and even human embryos can be frozen indefinitely yet remain viable; thereby, their vital processes are suspended. In time, it may become possible to freeze adult human beings. If so, it seems plausible to say that we could remain in existence indefinitely, in suspended animation. However, it remains controversial whether an organism is alive while in suspended animation.

The possibility of resurrection, raised by the third question, can seem straightforward: If we take a car completely apart, it ceases to exist; but if we reassemble the parts, the car's existence continues. Why should the same not be true of us? According to some theorists, it would. Typically, the molecules of which we are made disperse at death, but if they were later reassembled, just as they were when we died, our existence would continue.

There may be another way to survive death: Perhaps, contrary to appearances, we are immortal souls, as Descartes claimed—things distinct from, and capable of existing separately from, the bodies we inhabit. While the coherence of this suggestion has been challenged, the chief objection to it is that there is no evidence for the existence of souls.

Steven Luper

See also Death in the Social Sciences; Personal Identity and Trauma

Further Readings

Feldman, F. (2000). The termination thesis. *Midwest Studies in Philosophy, 24,* 98–115.

Hasker, W. (1999). *The emergent self*. Ithaca, NY: Cornell University Press.

Luper, S. (2009). *The philosophy of death*. Cambridge, England: Cambridge University Press.

McMahan, J. (2002). *The ethics of killing*. Oxford, England: Oxford University Press.

Olson, E. (2007). *What are we? A study in personal ontology.* Oxford, England: Oxford University Press.

Parfit, D. (1984). *Reasons and persons*. Oxford, England: Clarendon Press.

Snowdon, P. (1990). Persons, animals, and ourselves. In C. Gill (Ed.), *The person and the human mind: Issues in ancient and modern philosophy* (pp. 83–107). Oxford, England: Oxford University Press.

van Inwagen, P. (1978). The possibility of resurrection. *International Journal for the Philosophy of Religion, 9,* 114–121.

van Inwagen, P. (1990). *Material beings*. Ithaca, NY: Cornell University Press.

Death in the Social Sciences

The interdisciplinary study of the concept of death has paralleled the interest this topic has evoked among the general population. Although some early interest is known to have existed during the early 1920s, a much greater interest emerges in the aftermath of World War II and, more particularly, during the latter half of the 20th century and well into the second decade of the 21st century. Since the 1950s, interest in identifying cultural meanings and the social correlates of death has become an established area of inquiry among academics and practitioners alike.

This entry provides an overview of the concept of death, which is shown to be a continuous, evolving area of interest within the social sciences, not unlike that found in the behavioral and biological sciences.

Definition of Death and Application of the Death Concept

Defined by *Webster's New World Dictionary* as the "act or fact of dying," death is more than a biological concept; indeed, death has legal as well as medical consequences. In the former instance, death occurs when bodily functions close down and vital signs cease to be recorded. This legal definition is promoted as brain death, or the cessation of brain function. In the latter, medical perspective, death is described as the shutdown of circulatory and respiratory functions and of all the functions of the brain. Medically, death is the irreversible end of or absence of life; the end of life is the fact of brain death. But whether scholarly or applied, from a social science perspective, death is often defined through *metaphors,* demonstrated, for example, through the meaning attributed to cultural artifacts such as the law and medicine.

The historical and contemporary social-scientific analyses of mortality emerge from the four official categories of death: namely, *natural death, accidental death, homicide,* and *suicide*. Instances of death other than natural death, certifiable by an attending physician, must be documented by either a medical pathologist or a physician coroner. The need for such expert certification has long been a matter of contention in the science research literature as well as social policy assessment of this issue.

During the past 70 years, death has been found to be a *fluid* concept, as is noted in the vast social science literature. Indeed, death may be one of the most researched interdisciplinary topics, and for this reason, the analyses of specific types of biological death may be motivated by political and economic interests, especially within the medical research arena, wherein lies a great dependence upon government research funding.

The Early Emphases

The study of culture offers insights into the symbolic rituals that serve the need of a society to communicate the meaning of dying and death. The early emphasis on mortality focused on cultural practices specific to death, and this led to the establishment of subareas of interest within individual social sciences. Later, this research would take on an identity of its own, namely interdisciplinary death studies.

The data-gathering effort of the early-19th-century moral statisticians serves as the precursor for the social-scientific study of death. First published in 1897, Émile Durkheim's classic study *Suicide: A Study in Sociology* portrayed the correlates and causes of suicide, thereby setting the path for future concern with methodology in social science. At the same time, the cultural anthropological analyses of death-related cultural values, social rituals, and ceremonial rites also were to prove important to the evolution of this concept. Moreover, the development of probability theory and statistics in Europe during the latter part of the 19th century and the adoption of these methods by U.S. scholars were to have a profound effect on the study of death, especially in the application of actuarial tables to age-specific calculation of life insurance policies.

The 1920s brought a significant gain in the role of death-related research and the beginning of the death awareness movement. Three decades later, accounts of death camps and the reported annihilation of millions of civilians focused the public's attention on the reality of death, a reality that was introduced into homes and vividly portrayed through television. Finally, the writings of the displaced European intellectuals in the aftermath of World War II enhanced scholarly interest in the social aspects of death.

Death and Dying Movement

During the 1960s, 1970s, and 1980s, the mortality concept became a focal point of the death and dying movement. This movement transcended specific discipline boundaries to spawn a large and varied cultural and social structural dying and death literature. Each decade was to witness a few publications of special importance among the general population and a community of scholars.

Important examples include the 1959 work edited by Herman Feifel and titled *The Meaning of Death,* a book that arguably established the requisite foundation for the study of death; the exposé of the funeral industry, especially the high cost of funerals noted in the popular report *The American Way of Death,*

published in 1963; and, 10 years later, an assessment of the public perspective in *The Denial of Death,* a book by Ernest Becker. Furthermore, these research efforts included the assessment of the respective roles of the general physician and elected officials/medical examiners, such as coroners, and the evolving role of highly trained medical pathologists.

Exponential Growth of Higher Education and a New Academic Discipline

Early thematic social science books and edited anthologies underscore the topics that capture the attention of the public and members of academe. As the sociologist Robert Fulton declared in a personal communiqué, the 1960s was (to be) the decade for the study of one aspect of death, namely, suicide.

Soon, classes in death and dying were introduced into the growing North American and Western European college and university curricula. The emphasis that began in the late 1950s became more accepted throughout the following decades, as the study of death and dying became an important component of university and college course offerings. By the mid-1970s, interdisciplinary academic departments were formed, and thousands of classes that related to death and dying and other aspects of the human condition were offered.

The cross-cultural research proliferated to include topics such as bereavement and grieving; the respective roles of the dying patient and loved ones; the effects of tobacco; intimacy/sexuality; methods of death; the demography of death; the death care industry; religion and death; funerals and funeral facilities; the meaning of death, hospice, suicide, criminal homicide, drug-related deaths, art of death, fear of death, the role of religion; and age-related views of death. During the next several decades, other areas of research and clinical interest emerged.

The Application of New Knowledge

More recent insights into death and the human condition extend far beyond the earlier considerations. Noteworthy are euthanasia, rational suicide, reevaluations of fear and the social recognition of death, religion and death perceptions, and esoteric topics such as cryonics. Included also in this assessment are the laws relating to funeral and burial practices.

Important research also focuses on child and adolescent bereavement and short- and long-term consequences of the death of a parent. Other noteworthy areas include issues relating to the dying process and topics focusing on the living. All enhance the need to understand the complex litigious and social issues that heretofore were taboo topics, such as medical mistakes and wrongful death. Additional topics include pet cemeteries, drive-through funerals, and perpetual memorials.

In sum, death studies emerged as an interdisciplinary entity, and social scientists continue their quest to learn more about the macro- and microsocial organization of death-related activity. It is expected that future death-related insights will be gleaned from the multicultural aspects of death. Among these areas of interest will be a continuing focus on organ transplants; the funeral process; cyberspace memorialization; cremation; social control and death; the demographics of death; quality of life among an increasing aged, worldwide population; the respective changed roles of the clergy, funeral directors, and grief counselors; hospice; and a changed infrastructure for the expected incapacity of a growing number of the terminally ill.

Questions continue to be raised. These include issues relating to organ transplants, death notification, the need to conduct autopsies, community responses to death, and survival in the aftermath of natural disasters, such as tornados, hurricanes, and flooding, and megadeaths resulting from disasters of human origin.

Dennis L. Peck

See also Death and Immortality, Philosophical Perspectives; Durkheim's Philosophy of Social Science; Social Practices; Symbolism; Transhumanism and Human Enhancement

Further Readings

Bryant, C., & Peck, D. L. (Eds.). (2009). *Encyclopedia of death and the human experience.* Thousand Oaks, CA: Sage.

Kearl, M. (2009). Death, sociological perspectives. In C. Bryant & D. L. Peck (Eds.), *Encyclopedia of death and the human experience* (pp. 291–296). Thousand Oaks, CA: Sage.

Osterweis, M., Solomon, F., & Green, M. (Eds.). (1984). *Bereavement: Reactions, consequences, and care.* Washington, DC: National Academies Press.

Pine, V. R. (1977). A socio-historical portrait of death education. *Death Education, 1,* 57–84.

Debunking Social Science

To say that a subject has been "debunked" is to imply that despite its pretensions toward fulfilling some grandiose standard, it has fallen so radically short that its claims are revealed as not merely false but in some sense fraudulent. When one speaks of debunking social science, the customary argument is between those who contend that social science is on par with natural science—in theory if not quite yet in practice—and those who assert that the social sciences are just too soft in their methodology and too shot through with ideology to be taken seriously as sciences.

It is important to realize that apart from the question of whether the social sciences have actually succeeded in meeting scientific standards, there is a prior debate about whether they should even be trying to be scientific. One can, that is, attempt to debunk not only the idea of whether the social sciences have succeeded in fulfilling the scientific ideal but also whether the emulation of scientific standards is an appropriate ideal in the study of human behavior in the first place.

The claim that social science should (and could) meet scientific standards goes back at least to the Enlightenment, when the fruits of the Scientific Revolution began to shape our understanding of the natural world and held out the prospect that similar success might be enjoyed by those inquiries that emulated scientific methodology. If the foundation for social science is the desire to get to the truth about what causes human action, some have felt, what better way to get there than to embrace the empirical and experimental method that has given us the best basis for objective knowledge in human history?

Others have held, however, that this is a false ideal and that the problem is with the standard itself, since not even the natural sciences can live up to the logic of science that has been limned by philosophers. The idea here is that one needs to debunk not only social science but natural science as well, given recent philosophical work that purport to show that rampant sexism and racism pervade the allegedly "objective" investigation of the natural world, as well as other speculations having to do with the relativity and social construction of human knowledge.

Many have disagreed with such claims, however, and have held that they are belied not only by the internal problems within postmodern theory but also by the practical success of natural science. On this view, the problem is more specific to the social sciences. In the natural sciences, some have argued, we have done the best job that we can of eschewing ideology and making empirical compensations for human fallibility. But in the social sciences, the underlying problem exists not just in the alleged foibles of the investigator but in the very subject under inquiry. Humans are conscious; they may even have free will. As such, their behavior is subject to such a complicated set of self-referring influences and spontaneous causes that any piece of behavior must be accepted as open-ended. Science is just not possible, they claim, when one has a subject matter that is capable of such constant and unpredictable change.

There inevitably follows the question of whether social science *can* be more scientific. Until recently, many people on both sides of the debate about *whether* social science should be more scientific believed that it had done a poor job of living up to the standards set by natural-scientific investigation. Over and above the theoretical problems of how to contend with explanation in an open system, many felt that there are intractable practical problems in devising a better methodology for the social sciences, due to the corrupting influence of political ideology, cherry picking favorable results, failure to change one's theory in the face of falsifying instances, and ethical barriers to human experimentation, which had conspired to hobble the success of human inquiry. Recent work in the field of behavioral economics, however, has shown that it is possible to use better methodological techniques to contend with such problems, even if one cannot use them to decide questions such as free will.

Whether it is possible for the social sciences to build on this success and continue to make human behavior more predictable (and explicable) in the rest of the social sciences, however, must be seen as separate from the truth or falsity of any single theory. For one must recognize that all scientific theories—even well-accepted ones from natural science—are subject to empirical refutation. Isaac Newton's theory of gravity, for instance, is technically speaking false, but few would be so rash as to dismiss it as unscientific. To say that a theory is refuted by the evidence is not, therefore, to say that it has been debunked, for to debunk a claim is to show that it never had any business pretending to

the status that it had claimed for itself, which is different from saying (as in "refutation") that although the theory might have been true if the world had turned out to be a different place, it simply wasn't true given the way the world happens to be.

In this way, as more and more social-scientific theories are assessed not on their conformity with some abstract philosophical ideal but rather by their empirical success, one may discover that it is possible for the study of human behavior to hew more closely to the standards of scientific explanation. For even if this standard is still subject to the claim that it should be debunked, it is significant to note that this would amount to more than a mere debunking of social science, as it would require us also to recognize that all scientific theories—whether they are true or not—are trying to live up to the same ideal of finding true causal explanations.

Lee McIntyre

See also Causal Explanation, in Philosophy of Science; Explanation, Theories of; Pseudoscience; Science and Ideology; Scientific Method

Further Readings

Flyvbjerg, B. (2001). *Making social science matter.* Cambridge, England: Cambridge University Press.

Little, D. (1990). *Varieties of social explanation.* Boulder, CO: Westview Press.

McIntyre, L. (1996). *Laws and explanation in the social sciences: Defending a science of human behavior.* Boulder, CO: Westview Press.

Rosenberg, A. (1988). *The philosophy of social science.* Boulder, CO: Westview Press.

Decision Theory

Decision Theory is the theoretical study of decision making. Sometimes the term *Decision Theory* is used in a broad sense. It then includes game theory and Social Choice Theory. However, in many contexts, Decision Theory refers to the theoretical study of decisions taken by a single agent in which the outcome is independent of what other agents do. This entry follows that convention and discusses Decision Theory from this more restricted point of view, which is widely accepted in the literature.

Theories of decision making are commonly divided into *normative* theories and *descriptive* ones. The aim of normative decision theories is to formulate a coherent view about what we ought to do given the information about the world available to us at the point of making the decision. Descriptive decision theories on the other hand seek to describe—that is, explain and predict—how people actually make decisions. Numerous empirical studies have shown that people frequently violate the recommendations of the most well-established normative theories.

Maximizing Expected Value

The history of normative Decision Theory began in 1654, as Blaise Pascal and Pierre de Fermat developed the concept of probability, which they used for analyzing which bets or gambles a rational player would accept. A few years later, in a book by Antoine Arnauld and Pierre Nicole (1662/1996) commonly known as the *Port Royal Logic,* we find the first precise formulation of what is still the most widely accepted decision rule among decision theorists, namely, the principle of maximizing expected value:

> In order to judge of what we ought to do in order to obtain a good and to avoid an evil, it is necessary to consider not only the good and evil in themselves, but also the probability of their happening and not happening, and to regard geometrically the proportion which all these things have, taken together. (p. 274)

Let us illustrate this decision rule in a concrete example. Imagine, for instance, that you wish to figure out whether it is worth paying £30 for travel insurance that would cover your medical expenses if you were to get ill when visiting a foreign country. Whether it is rational to buy this insurance arguably depends on how likely it is that you will get ill, as well as on the cost of medical treatment in the country you plan to visit. Suppose that you find out, in one way or another, that one person per 1,000 travelers gets ill in the country you plan to visit and that the typical cost for medical treatment is £5,000. This means that if you pay £30 you can avoid a probability of .001 of having to pay £5,000. Would it be rational to take out such an insurance policy?

The principle of maximizing expected value prescribes that an act is rational if and only if its

expected monetary value is at least as high as that of every other alternative act. The expected value of an act is defined as the sum total of the values of all possible outcomes, weighted by the probability of each outcome. In our example, there are two alternative acts: buy insurance or do not. Each act will result in exactly one of two possible outcomes: You either get ill or you do not. This entails that the expected monetary value of buying the insurance is –£30 + .001 × £5,000 = –£25, whereas the expected monetary value of not buying it is .001 × £5,000 = –£5. It therefore seems that we have to conclude that it would not be rational to buy the insurance.

The Concept of Utility

Throughout the history of Decision Theory, scholars have discussed how to exactly formulate the principle of maximizing expected value. In 1738, Daniel Bernoulli pointed out that it is not an act's expected *monetary* value that matters but rather what we nowadays call its expected *utility.* The concept of utility is much broader and includes all aspects of an outcome that decision makers may care about, such as the fact that one would perhaps feel safer if one buys travel insurance before visiting a foreign country. Moreover, as pointed out already by Bernoulli, many people have a decreasing marginal utility for money. This means that a large amount of money is not worth twice as much to you as half of the large amount. The first million you earn typically matters more to you than the second. Therefore, decision theorists seeking to apply the principle of maximizing expected value need to find ways of replacing the monetary values of all the possible outcomes with some linear representation of value, and the concept of utility is designed to do exactly that. By definition, the difference in utility between 2 million units of utility and 1 million is exactly as large as the difference in utility between 1 million and 0 units.

Modern research in Decision Theory has been dominated by attempts to construct axiomatic arguments for the principle of maximizing expected utility. The agenda for this endeavor was set by Frank Ramsey's paper "Truth and Probability," written in 1926 but published posthumously in 1931. In this paper, Ramsey proposed a set of eight axioms for how rational decision makers ought to choose among uncertain prospects. He noted that a decision maker who acts in ways that are compatible with his axioms will always perform acts that are *compatible* with the principle of maximizing expected utility, by *implicitly* assigning numerical probabilities and utilities to outcomes. However, Ramsey did not claim that the decision maker's choices were *actually triggered* by these implicit probabilities and utilities.

Another important point of departure for modern work in Decision Theory is John von Neumann and Oscar Morgenstern's book *Theory of Games and Economic Behavior,* in particular the second edition published in 1947. In that edition, they presented a set of axioms for how rational decision makers ought to choose among lotteries. The term *lottery* is used here in a broad sense. Imagine, for example, that you prefer a BMW to a Volvo, and a Volvo to a Ford, and that you are indifferent between getting a Volvo for certain and a 50:50 chance of getting a BMW or a Ford. Then, a BMW is worth 1, a Volvo 0.5, and a Ford 0 (or any positive linear transformation of these numbers) on your personal utility scale. Why? Because given that you evaluate lotteries in coherent ways, your preferences have revealed how much value (utility) you place on the three cars. This means that your utility function for cars is *defined* in terms of preferences over lotteries.

Von Neumann and Morgenstern showed that every decision maker who acts in accordance with their axioms implicitly behaves in accordance with the principle of maximizing expected utility. This means that we can assign numerical utilities to outcomes that help us explain and predict how the decision maker will behave in the future. The main difference between von Neumann and Morgenstern's work and that of Ramsey is that Ramsey also presented a novel (subjective) theory of probability. In his 1954 book *The Foundations of Statistics,* Leonard Savage presented a somewhat different but also very influential axiomatic analysis of the principle of maximizing expected utility. Many of Savage's ideas can be found in Ramsey's paper.

Bayesian Decision Theory

Explicit discussions of which act to perform in a given situation are only a minor part of what Decision Theory is about. The key issue in the contemporary debate is, rather, how basic concepts such as preference, utility, probability, and expected utility should be defined and interpreted. For many years, the dominating view has been something called

Bayesian Decision Theory—after Thomas Bayes, the English mathematician and Presbyterian minister (1701–1761). Bayesians claim that probabilities are subjective degrees of belief that should be updated in light of new evidence by applying Bayes's theorem. This theorem, also known as the *reversed probability law,* says in its simplest form that the probability of event A given event B is equal to the probability of A multiplied by the probability of B given A, divided by the probability of B. (Note that Bayes's theorem follows deductively from the axioms of the probability calculus, so its truth is not open to any fruitful criticism or disagreement, but its scope and relevance are.) Bayesians furthermore maintain that probabilities are subjective degrees that should be defined, together with utilities, in terms of preferences over risky gambles, which can in turn be used for indirectly justifying the principle of maximizing expected utility.

The assumption that preferences are revealed in choices can be criticized. One objection is that we sometimes seem to *be,* and *feel,* uncertain about our own preferences. Suppose, for instance, that Anne is about to vote in an upcoming election. She knows for certain that she prefers both the Liberal Party candidate (L) and the Conservative Party candidate (C) over the candidate representing the Communist Party. But she is uncertain about her preference between the Liberal candidate and the Conservative candidate. In a traditional theory of preference, all one can say about this situation is that Anne's preference ordering is incomplete; that is, all of the following are false: L is preferred to C, C is preferred to L, and L and C are indifferent. However, such an analysis overlooks significant aspects. Suppose, for example, that Anne is almost sure that she prefers the Liberal candidate and feels that she would vote for that candidate 9 times out of 10 (not changing any of her mental dispositions). If this is the case, it is not reasonable to maintain that Anne, at each point of time, has a certain preference and that this preference wobbles between L and C; the concept of uncertain preference models the situation better.

Another objection to Bayesian Decision Theory is that from a linguistic point of view it seems rather odd to say that the *meaning* of utility has something to do with preferences among uncertain prospects. Even a decision maker who believes that she lives in a deterministic world—that is, a world in which every act has a well-determined outcome—can meaningfully say that the utility of some certain outcome is higher than the utility of another certain outcome. Although orthodox Bayesians never assign probability 1 or 0 to any outcomes, it seems that this inability to account for our considered intuitions in this type of case is a serious problem.

Objections to the Expected Utility Principle

A lot of recent work in Decision Theory aims to cast doubt on the principle of maximizing expected utility, as both a normative principle as well as a descriptive claim. The starting point for much of this criticism is a famous example discussed by Maurice Allais in the 1950s. Consider the following lotteries, in which exactly one winning ticket will be drawn at random.

Empirical studies by Daniel Kahneman and Amos Tversky and others show that in a choice between Lottery 1 and Lottery 2 about 80% of ordinary people prefer Lottery 1. From a normative point of view, this seems very reasonable, since Lottery 1 gives the decision maker £10 million for sure. Moreover, the same empirical studies also show that in a choice between Lottery 3 and Lottery 4 about 80% feel that it is rational sense to trade a ten-in-hundred chance of getting £50 million against a 1 in 100 risk of getting nothing, and consequently, they prefer Lottery 4. However, the point of the example is that no matter what utility one assigns to money, the principle of maximizing expected utility recommends that the decision maker prefer Lottery 1 to Lottery 2 *if and only if* Lottery 3 is preferred to Lottery 4. There is simply no utility function such that the principle of maximizing utility is consistent with a preference for Lottery 1 to Lottery 2 *and* a preference for Lottery 4 to Lottery 3. In order to prove this, we calculate the *difference* in expected utility between the two pairs of lotteries. The probability that Ticket 1 will be drawn is .01, and the probability that one of the tickets numbered 2 to 11 will be drawn is .1; hence, the probability that one

	Ticket No. 1	*Ticket No. 2–11*	*Ticket No. 12–100*
Lottery 1	£10 million	£10 million	£10 million
Lottery 2	£0	£50 million	£10 million
Lottery 3	£10 million	£10 million	£0
Lottery 4	£0	£50 million	£0

of the tickets numbered 12 to 100 will be drawn is .89. We, thus, have the following equations:

$$u(\mathrm{L1}) - (u(\mathrm{L2}) = u(1\mathrm{M}) - [.01u(0\mathrm{M}) + .1u(5\mathrm{M}) + 0.89u(1\mathrm{M})] = .11u(1\mathrm{M}) - [.01u(0) + .1u(5\mathrm{M})] \quad (1)$$

$$u(\mathrm{L3}) - u(\mathrm{L4}) = [.11u(1\mathrm{M}) + .89u(0)] - [.9u(0\mathrm{M}) + .1u(5\mathrm{M})] = .11u(1\mathrm{M}) - [.01u(0) + .1u(5\mathrm{M})], \quad (2)$$

where L represents Lottery and M represents million.

By looking at Equations 1 and 2, we see that the difference in expected utility between L1 and L2 is the same as the difference between L3 and L4. We can therefore conclude that no matter what the decision maker's utility of money is, it is impossible to simultaneously prefer L1 to L2 *and* to prefer L4 to L3 without violating the expected utility principle. However, since many people still think it would be rational to stick to the preference pattern described above, it seems that there is something wrong with the expected utility principle.

Decisions Under Ignorance

Scholars who reject subjective theories of probabilities (according to which probabilities are subjective degrees of belief) think that there are decision problems in which it would be meaningless to ascribe probabilities to outcomes. Imagine, for instance, that you are thinking about getting married to your fiancé. What is the (objective) probability that your particular marriage would be a happy one? It would obviously make little sense to study other marriages, since every marriage is unique. Decision theorists call this type of decision, in which you do not know the probabilities of the possible outcomes, "decisions, under *ignorance*" (or under *uncertainty*). Two of the most influential decision rules for dealing with decisions under ignorance are the *maximin* rule and the *principle of insufficient reason*. The maximin rule was famously discussed by John Rawls in his work on justice. The key idea behind this rule is that the decision maker should *maximize* the *minimal* value obtainable with each alternative act. If the worst possible outcome of getting married is worse than the worst possible outcome of remaining unmarried, then it is rational to refuse marriage.

The principle of insufficient reason, as proposed by Bernoulli and Pierre-Simon Laplace, urges us to transform our ignorance about the relevant probabilities into a decision under risk by assigning *equal* probabilities to all possible outcomes of an act. That is, if one has *no* reason to think that one outcome is more probable than another, then all outcomes should be assigned equal probability. It is sometimes objected that it seems completely arbitrary to infer that all outcomes are equally probable if one has no reason to think that one outcome is more probable than another. In such a case, it seems strange to conclude anything at all about probabilities.

Some Contemporary Problems in Decision Theory

An important debate in contemporary Decision Theory concerns the controversy over *causal* and *evidential* Decision Theory. In many decision problems, the decision maker's beliefs about causal processes play a significant role. However, it is very difficult to give a convincing explanation of what role beliefs about causal process should be allowed to play, as illustrated in an example known as the *smoking lesion*. We know that smoking is strongly correlated with lung cancer; but imagine that in some remote world, lung cancer and the urge to smoke have a common, perhaps genetic, cause. In that world, smoking does not cause cancer; the two are rather effects of a common cause. Now imagine that you prefer smoking without cancer to not smoking without cancer and that you also prefer smoking with cancer to not smoking with cancer. Given these assumptions, would it be rational to smoke? Although you may have a strong intuition about what it would be rational to do in this case, there are other, structurally similar cases that are much harder to analyze.

Another major subfield of modern Decision Theory is multi-attribute Decision Theory. While in a single-attribute approach all outcomes are compared on a single utility scale, multi-attribute approaches seek to evaluate the outcomes along the scale deemed to be the most appropriate for the type of value at stake. Imagine, for instance, that we have to decide whether it is worth spending £10 million on a program that aims to increase road safety by saving one extra statistical life per year or whether it would be better to spend the money on reducing the number of nonlethal accidents aboard fishing

vessels. Many people think that such direct comparisons between the value of a human life, money, and other types of accidents are impossible, meaning that these entities are incomparable values. Now, multi-attribute Decision Theory seeks to avoid the criticism that, for example, money and human welfare are incomparable values by relaxing the assumption that all outcomes have to be compared on a common scale. In multi-attribute Decision Theory, each type of value is measured in the unit considered to be the most suitable for that value. While money appears to be the right unit to use for measuring financial costs, the number of quality-adjusted life-years might be the unit we should use for measuring human welfare. The total value of an alternative act is then calculated by aggregating the different values into an overall ranking of the available alternatives.

Martin Peterson

See also Allais Paradox; Bayesianism, Recent Uses of; Collective Rationality; Cost–Benefit Analysis; Game-Theoretic Modeling; Preference; Rational Choice and Political Science; Rational Expectations; Rationality and Social Explanation; Risk; Social Choice Theory

Further Readings

Allais, M. (1953). Le Comportement de l'homme rationnel devant le risque: Critique des postulates et axiomes de l'ecole Américaine [Rational man's behavior in the presence of risk: Critique of the postulates and axioms of the American school]. *Econometrica, 21,* 503–546.

Arnauld, A., & Nicole, P. (1996). *Logic or the art of thinking* (5th ed., J. V. Buroker, Ed. & Trans.). Cambridge, England: Cambridge University Press. (Original work published 1662)

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica, 47,* 263–291.

Nozick, R. (1969). Newcomb's problem and two principles of choice. In N. Rescher (Eds.), *Essays in honor of Carl G. Hempel* (pp. 114–146). Dordrecht, Netherlands: Reidel.

Pascal, B. (1910). *Pensées* [Thoughts] (F. W. Trotter, Trans.). New York, NY: P. F. Collier.

Peterson, M. (2009). *An introduction to decision theory.* Cambridge, England: Cambridge University Press.

Ramsey, F. P. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays* (pp. 156–198). London, England: Kegan Paul.

Savage, L. J. (1972). *The foundations of statistics* (2nd ed.). New York, NY: Dover.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

DEDUCTION

In ordinary language, the term *deduction* vaguely refers to a kind of activity in which we all are involved when we try to solve theoretical or practical problems: drawing conclusions from given premises. In logical theory, the term has a more precise meaning and refers to the special case in which the conclusion is *indisputably* justified on the basis of the premises. Deductive reasoning plays a pervasive role in philosophical and scientific practice as well as in everyday problem solving and decision making, and the ability of a "rational agent" to perform deductive inferences (no matter how complex) is an idealized assumption of some economical, political, and ethical theories.

This entry offers an informal overview of the general concept of deduction. A rigorous treatment is the subject of formal logic.

From a linguistic viewpoint, an *inference* consists of a list of sentences (premises) followed by words like *hence, thus,* or *therefore* and then by another sentence (the "conclusion"). Consider the following two examples:

A:	B:
1. All ravens are black.	1. All ravens observed so far are black.
2. Mr. Poe is a raven.	2. Mr. Poe is a raven.
Therefore: Mr. Poe is black.	Therefore: Mr. Poe is black.

In both examples, the premises provide some justification for asserting the conclusion. However, in Inference A, the justification is as strong as it gets: No rational agent can fail to recognize that its conclusion must *certainly* be true in all situations in which the premises are true. This is what we mean by saying that A is *deductively sound*. In contrast, it is perfectly conceivable that the premises of B are true while its conclusion is false. So B is not deductively sound and (much) harder to justify. According to some authors—most eminently Karl

Popper in his *Logic of Scientific Discovery*—it cannot be justified at all. According to others—most eminently Bertrand Russell in his *The Problems of Philosophy*—it may be justified by a general "principle of induction" and thereby regarded as *inductively* sound.

In order to recognize that A is sound, we do not need to understand the meaning of "raven" and "black," or the denotation of "Mr. Poe." Indeed, all these words could be substituted by other words, such as *whales, white,* and *Moby Dick,* without affecting the soundness of the inference. However, substituting "all" with "some" would yield an obviously unsound inference. Hence, the soundness of A depends only on the meaning of "all." A word whose meaning is essential for recognizing the soundness of an inference is called a *logical word* or *logical constant*. The other words are called *extralogical* and can be replaced by schematic letters. So the general form of A is as follows:

A*:

1. All *U* are *V*.

2. *x* is *U*.

Therefore: *x* is *V*.

A schematic inference such as A* is called an *inference rule*. The soundness of all its instances can be immediately recognized by any agent who understands the meaning of "all." (Indeed, some authors maintain that the meaning of a logical word can be completely defined by exhibiting simple inference rules, such as A*, showing how the logical word can be legitimately used in inference.) So, recognizing that an inference is deductively sound, in the simplest cases, is just part of our linguistic competence. This is why deductive reasoning has often been described as "analytic," and it has been maintained, especially by logical neopositivists, that it conveys no new information. (This claim, however, is intuitively implausible, and Jaakko Hintikka has called it a true "scandal of deduction.")

When the inference is not an instance of any basic inference rule, and so its soundness is not immediately obvious, we may still be able to fill the gap between premises and conclusion by constructing a *deduction*—that is, a sequence of "steps" each of which is an instance of some obviously sound rule. If we fail, we can try to show that the inference is *not* sound by exhibiting a counterexample, namely, a possible situation in which the premises are true but the conclusion is not.

This is, however, only the skeleton of a suitable notion of deduction. Specific notions are obtained by specifying which words are to be regarded as logical and the basic inference rules that capture their meaning. In classical logic—whose modern roots can be found in the works of George Boole, Gottlob Frege, Ludwig Wittgenstein, Bertrand Russell, David Hilbert, and Alfred Tarski—the logical words are the Boolean operators "and," "or" (in its inclusive sense), and "not," plus the quantifiers "all" and "some." These logical words allow also for an approximation of the conditional "if *P*, then *Q*"—defined as "not-(*P* and not-*Q*)"—which works reasonably well in mathematical contexts but is quite removed from ordinary usage in most of the others. (This meaning of the conditional operator was proposed by Philo of Megara in antiquity and was already criticized by the Stoic Chrysippus, who proposed an alternative definition that can be related to C. I. Lewis's *strict conditional*.)

The classical meaning of the logical words is intertwined with the Aristotelian "Principle of Bivalence," according to which a sentence is determinately either true or false, independently of our epistemic means for recognizing its truth or falsity. This hidden metaphysical assumption provided the *intuitionists* a strong motivation for rejecting the soundness of all inferences depending on it—such as the rule of proof *ex absurdo,* by which means one shows that the conclusion follows from the premises by showing that its negation is inconsistent with them—on the grounds of their mathematical anti-realism. Other nonclassical logics that have been proposed in the 20th century depart from the classical paradigm either because, like intuitionistic logic, they reject some classical inferences as unsound (this is the case of Relevance Logic) or because they extend the stock of logical words (this is the case of modal and epistemic logics). As a result of recent developments, especially in the areas of nonmonotonic, substructural, paraconsistent, and quantum logics, the classical paradigm has been replaced by a proliferation of logical theories that are perceived as fulfilling different application needs rather than competing with each other, and the very concept of deduction has been considerably stretched far beyond its standard meaning discussed in this entry.

Marcello D'Agostino

See also A Priori and A Posteriori; Analytic/Synthetic Distinction; Hypothetico-Deductivism; Induction and Confirmation; Logical Positivism/Logical Empiricism

Further Readings

Gabbay, D. M., & Guenthner, F. (Eds.). (2001). *Handbook of philosophical logic* (Vols. 1–16, 2nd ed.). Berlin, Germany: Springer.

Hintikka, J. (1972). *Logic, language-games and information*. Oxford, England: Oxford University Press.

Popper, K. (1959). *Logic of scientific discovery*. New York, NY: Routledge. (Original work published 1934 in German as *Logik der Forschung*)

Quine, W. V. O (1982). *Methods of logic* (4th ed.). Cambridge, MA: Harvard University Press.

Russell, B. (1912). *The problems of philosophy*. London, England: Home University Library.

Tennant, N. (1990). *Natural logic* (2nd ed.). Edinburgh, Scotland: Edinburgh University Press.

Deontic Logic and Agency

Deontic Logic is the branch of formal logic that aims to model the reasoning of agents who are subject to obligations, prohibitions, and permissions relative to a normative system. *Agency* is the philosophical term for an agent's capacity to influence the possible courses of history our world can take. For a thorough understanding of both, Deontic Logic and theories of agency are often studied in combination.

Obligations, permissions, and prohibitions are always defined relative to a normative system. These normative systems come in various forms. They can be private and determine an agent's response to personal attitudes, such as when an agent believes she has the duty to do something with her talents. Or they may live in what John Searle calls *social reality,* such as when agents have to adhere to the norms, conventions, or unwritten rules of conduct their membership in a society, culture, or religion brings with it. Or normative systems may be formalized and made explicit in law books, contracts, product user guides, cookbooks, and so on. Deontic Logic aims to mathematically model the reasoning with deontic modalities explicitly or implicitly defined relative to such normative systems.

Although there were earlier attempts at formalizing deontic reasoning, it is generally agreed upon that modern Deontic Logic started with Georg Henrik von Wright's article "Deontic Logic" in 1951. In this work, the three central deontic modalities—obligation, prohibition, and permission—are formalized as the operators of what later came to be known as *Standard Deontic Logic* (SDL), a formalism that logicians nowadays classify as a modal logic of the type KD. Roughly, von Wright's SDL drew two main lines of criticism. The first concerned the fact that its logical operators could not faithfully model conditional obligations, such as "If you help, you have to tell that you will," and so-called contrary-to-duty obligations, such as "You have to help, but if you do not, you should not tell that you will." This led to the study of dyadic Deontic Logic operators for conditional obligations, with dyadic operators O(p|q) read as "Given that q, it is obliged that p." Based on the philosophical viewpoint that if (dyadic) deontic operators represent norms—in other words, think of them as the rules in a law book—they cannot bear truth-values (since they are prescriptive and not descriptive, i.e., they express an "ought" not an "is"), David Makinson transformed the dyadic approach to what he referred to as developing systems of *iterative detachment*. Together with the work of Leon van der Torre, this new approach led to the study of *input–output logics*.

The second type of criticism toward von Wright's SDL concerned the fact that it aims to study deontic modalities in isolation, while in any concrete example of deontic reasoning, modalities of time, action, intention, belief, and so on interfere. It was suggested then that the many counterexamples to the modeling power of SDL that were found—the so-called deontic paradoxes—could be accommodated by extending Deontic Logic with suitable operators for action, time, and knowledge.

In line with this second type of criticism, the concept of *agency* also was observed as missing from the conceptual repertoire of SDL. Agency, as we have seen, is the philosophical term for any agent's apparent capacity to influence the course of history our world can take. Philosophers distinguish here between events and actions. Events might occur without agentive involvement. Actions, though, are always the result of decisions by agents and of the agentive effort to exert these choices. That the phenomenon of agency provokes deep philosophical questions becomes evident right away if one realizes that prima facie it is not clear how to reconcile it

with either causal determinism or indeterminism—namely, with the question of whether human agents are free to choose their action, that is, whether they have free will or, in contrast, are subject to causes they cannot control.

The questions of agency and of Deontic Logic were brought together for the first time in the work of Stig Kanger and further developed by, among others, his student Ingmar Pörn. The intimate relation between normative reasoning and agency becomes clear if we look at the notion of excuse. An agent can be (partially) excused for violating a prohibition if there are reasons for not holding him (completely) responsible for what he did. And an agent can be (partially) excused for not meeting an obligation if there are reasons for not holding him (completely) responsible for what he refrained from doing. It might be, for instance, that another agent was complicit in the wrongdoing or that an act was forced upon an agent by some other agent or that an agent was influenced by other external circumstances beyond her control. Early attempts at formalizing the type of agency that in the context of a normative system and a violation implies culpability focused on the counterfactual condition that an agent can only be held responsible if she could have chosen or acted otherwise. In the early works of Kanger and Pörn, this counterfactual condition was left implicit in the abstract semantics of the agency operators of the logic. Later, Nuel Belnap and John Horty further developed this line of work by adding the branching-time dimension to the models and the language of the logic. This resulted in logics where the central modality for "seeing to it that" (*stit,* for short) equates agency with selecting certain histories (timelines) and excluding others. The formalism enabled them to be more precise about the counterfactual side conditions involved in deliberative agency. This line of work culminated in Horty's influential book on agency and Deontic Logic, which studies different forms of agency, such as individual agency, group agency, strategic agency, and so on, in interaction with different types of normative modalities, such as ought-to-be, ought-to-do, or conditional obligation.

Recent developments in the theory of Deontic Logic and agency have acknowledged the role of epistemic and motivational modalities in the kind of agency that can be classified as nonexcusable culpable conduct. For instance, we might want a logic to reflect that only intentional action can lead to wrongdoing. Or we might want to incorporate that an agent is only responsible for acts he knows he does. By extending the framework of *stit* logic with operators for "knowingly doing," Jan Broersen showed recently how we can formally characterize several levels of *mens rea* (guilty mind) found in the legal literature. To date, many questions in this area are still open. For instance, it is not clear how to formalize group agency of the type where agents collectively know what they are bringing about. Also, it is not clear yet how exactly examples of moral luck can be modeled. Finally, more research will have to be devoted to questions concerning the formalization of partial responsibility, attempt, unsuccessful action, and probabilistic action.

Jan Broersen

See also Action, Philosophical Theory of; Agency; Coalition Logic; Determinism; Free Will, Philosophical Conceptions of; Mathematical Models, Use in the Social Sciences; Modal Logic and Intentional Agency; Social Norms

Further Readings

Broersen, J. M. (2011). Deontic epistemic stit logic distinguishing modes of mens rea. *Journal of Applied Logic, 9*(2), 127–152.

Chisholm, R. M. (1963). Contrary-to-duty imperatives and deontic logic. *Analysis, 24,* 33–36.

Davidson, D. (2000). Agency. In *Essays on actions and events* (pp. 43–62). Oxford, England: Oxford University Press. (Reprinted from *Agent, action, and reason,* by A. Marras, R. Binkley, R. Bronaugh, Eds., 1971, Toronto, Ontario, Canada: University of Toronto Press.)

Horty, J. F. (2001). *Agency and deontic logic.* Oxford, England: Oxford University Press.

Horty, J. F., & Belnap, N. D. (1995). The deliberative stit: A study of action, omission, and obligation. *Journal of Philosophical Logic, 24*(6), 583–644.

Kanger, S. (1971). New foundations for ethical theory. In R. Hilpinen (Ed.), *Deontic logic: Introductory and systematic readings* (pp. 36–58). Dordrecht, Netherlands: Reidel.

Nagel, T. (1979). Moral luck. In *Mortal questions* (pp. 24–38). Cambridge, England: Cambridge University Press.

Pörn, I. (1977). *Action theory and social science: Some formal models.* Dordrecht, Netherlands: Reidel.

von Wright, G. H. (1951). Deontic logic. *Mind, 60,* 1–15.

Determinism

The entry reviews some of the core philosophical issues about determinism and forges the links with the social sciences.

Determinism is the philosophical notion not only that every event has a cause but also that every event is the inevitable result of a causal chain that stretches backward in time to infinity. If true, this would mean that there are no random, uncaused events. This notion is controversial enough when applied to the natural world, where it quickly leads to some of the most complex (and unsettled) debates in subatomic physics. But when the notion of determinism is applied to human events, it is perhaps even more controversial, since its truth here would seem to imply that human beings are not in control of their own destinies, because many have felt that determinism is incompatible with free will.

In social science, claims about determinism concern two matters: First, that every piece of human behavior must be caused (which is problematic enough in some circles, since it would seem to imply that a scientific study of human behavior is possible) and, second, that these causes originate outside the person whose behavior we are studying, since each cause was itself presumably the result of some earlier cause that led up to it. In tracing such a line of causation to its origin, we inevitably arrive at a point before the person's birth. And how can one be responsible for this?

All the plans, passions, decisions, and motivations of human life would therefore threaten to be mere epiphenomena; if we are not the author of the events in our lives, but merely the mechanism through which the inevitable chain of events determined before our birth plays itself out, then what is the point of human cognition or accountability? Is the point of human life merely to feel that our actions are meaningful, even if we do not in fact cause them?

Debates about determinism in the social sciences thus quickly involve us in questions about free will, predictability, and moral responsibility for our actions.

Free Will

Is human behavior part of an open system? Are we capable of choosing to do whatever we like? If so, one wonders why we keep seeing the same sorts of tragedies repeat themselves in human affairs over the ages. But if not, some have felt that it is even more of a tragedy because there is nothing that we can do to ameliorate human suffering. Whether human systems are open or not, it seems important to try to square our feeling of freedom with the reality that human behavior seems at times to be highly regular.

Free will cannot mean that human events are totally capricious, for if it did, we could not hope to understand them. Surely, one cannot have a science where there is no regularity. But it also seems true that even a subjective understanding of human events would fail if there were not some close correlation between our inner awareness and the behavior that we (and others) exhibit.

Perhaps, some have argued, such a subjective understanding is consistent with the idea that free will originates in the beliefs and desires that influence human action. Yes, these beliefs and desires may cause our actions, but they also presumably come from us, thus locating the causal origin of human behavior in a place over which we seem to have some degree of control. This idea, however, suffers from the reality that our beliefs and desires at times seem heavily influenced—perhaps unbeknownst to us—by outside events. Even if our beliefs and desires are ours, have we been caused to have them?

This raises the specter that there might be some degree of causal force behind human events, over which we have no control. Would this conflict with free will? Not necessarily, for some might claim here that we are merely ignorant of the reasons behind our actions, even if we are the author of them, and when we someday become enlightened (perhaps through the social sciences), we will be better able to change our behavior simply by changing the beliefs and desires that govern it.

Either way, we would be wise to embark on a more systematic understanding of the causes of human action. Whether our behavior is completely determined or we are, in fact, free but merely ignorant of the ways in which our beliefs and desires have been manipulated by outside forces, it seems important to have a good deal more foreknowledge of how particular decisions on our part can lead to certain outcomes.

Prediction

The prediction of human action has been a goal of the social sciences since their inception. Based on the successful model of Newtonian mechanics, which views prediction and explanation as flip sides

of the same coin, one wonders whether the same could be true for human behavior. Specifically, if human action is determined, there would seem to be, in principle, no barrier to predicting human action with the same degree of precision one might use to predict an eclipse. With deterministic systems, if we understand the causes, then we understand the outcome. But there is a problem, for the types of systems that one is able to predict in Newtonian physics are not conscious and do not intentionally try to defy the predictions that are made about their actions once they become aware of them.

The problem of "reflexive prediction" seems unique to human affairs. But it is not, since natural science does have examples whereby a prediction made about a system comes to foul the very thing that one is trying to predict. (Examples abound, from the mundane, the prediction of an avalanche in a loud voice, to the subtle, the prediction of the momentum of a subatomic particle once we have already discerned its position, as per Heisenberg's Uncertainty Principle.) But given the degree to which such self-awareness is embedded in human consciousness, understanding a prediction being made about us does seem to present a special challenge to forecasting human behavior. If the chairman of the Federal Reserve predicts a rally in the stock market, it may very well happen. Likewise, if one predicts to a chronically tardy friend that he will be late for a movie, one should not be surprised to find him tapping his foot on the sidewalk in defiance when we arrive to find him 5 minutes early. Whether the prediction is "self-fulfilling" or "self-defeating," the complexity of human affairs seems to all but guarantee that any prediction about human action will itself become a link in the causal chain that influences the action in question.

It is important to note, however, that this tendency does not necessarily defeat the determinism of human events, nor does it prove that we have free will. The very fact that humans are regular in their behavioral response to predictions seems significant and may suggest that predictions can at times be used to manipulate human action even when we are trying our best to act freely. Did we predict that our friend would be late precisely because we wanted him to arrive on time for once? But surely, the advocate for free will points out, if he had known this, he could have acted differently, but then so would we, and so on, ad infinitum.

Still, the underlying problem here is one that is endemic to predictions across both the natural and the social sciences, whether determinism is true or not, which is that prediction is practically difficult when dealing with complex systems. Even if human affairs are not actually "open" (as one assumes they would be if we had free will) but are in fact merely "complex" (as they might be if our behavior were determined by multiple factors), it might still prove impractical to predict human affairs. As one sees in meteorology, extremely small influences can be amplified and thus confound a prediction, even when a system is completely causally closed. Whether or not human behavior is, in fact, predictable, therefore, would seem a poor proxy for discerning whether it is determined.

Moral Responsibility

There is yet another area of contrast between human and natural events, and this is that in studying inanimate matter we do not face the question of moral responsibility. If human behavior is determined, some philosophers have argued, then we are not morally responsible for it. If something is outside our control, how can we be praised or blamed for it? Others have claimed, however, that this reasoning does not hold, for even if an event is determined there still might be grounds for assessing moral responsibility; note that we tolerate complete prediction as compatible with moral responsibility in other realms. The classic example is the existence of an omniscient God. If God sees everything, then He can predict our behavior. But if He also gave us free will, then presumably we are nonetheless responsible for our actions, even if God foresaw them. Such "compatibalist" arguments are the source of widespread debate among determinist philosophers.

But there is another side to this paradox that is highlighted by those who reject determinism. If determinism is not true, then there are at least some uncaused human events. But if they are not caused, then they are not caused by me, so how can I be held responsible for them? Here, some philosophers have constructed a causal model to rescue the notion of moral responsibility, even in the face of indeterminism, whereby one has just enough causation to seat the event in the "agent" who is making the decision, but not so much causation that the agent is "caused to cause" her decision. This model too remains a source of controversy among philosophers.

Thus, the notion of human responsibility would seem to hang either on the thread of the success of compatibalist arguments in the face of determinism

or the claim that human free will is the very seat of causal efficacy that lies behind human action.

Epistemological Limits

Should the absence of resolution in the debate over human determinism be a barrier to the social sciences? In particular, should it obstruct the claim that we may pursue a scientific study of human behavior?

It seems inconsistent to argue that the debate over determinism in social science must be resolved in order to move forward in the study of human affairs when we have not yet resolved the issue of determinism in natural science. If anomalous results, like the possibility of violating causal laws at the subatomic level, are possible in physics, then surely this gives us some latitude in the face of insistence that the issue of human determinism must be resolved before we can make progress in the social sciences? After all, there has been plenty of progress in the natural sciences, notwithstanding their own controversies over determinism—witness the continued dispute over quantum mechanics.

Whether determinism is true—either in the social sciences or in some more ultimate sense—the question of whether we can pursue an empirical program of study in the social sciences would seem to rest less on the metaphysical question of whether human action is completely determined than on the epistemological question of how we may come to know the causes that underlie human behavior, whatever be their origin.

Lee McIntyre

See also Causation in the Social Sciences; Causes Versus Reasons in Action Explanation; Complexity and the Social Sciences; Explanation Versus Understanding; Free Will in the Social Sciences

Further Readings

Campbell, J. K., O'Rourke, M., & Shier, D. (Eds.). (2004). *Freedom and determinism*. Cambridge: MIT Press.

Earman, J. (1986). *A primer on determinism*. Dordrecht, Netherlands: Reidel.

Goldman, A. (1968). Actions, predictions, and books of life. *American Philosophical Quarterly, 5,* 135–151.

Honderich, T. (2005). *On determinism and freedom*. Edinburgh, Scotland: Edinburgh University Press.

Machlup, F. (1978). If matter could talk. In *Methodology of economics and other social sciences* (pp. 309–332). New York, NY: Academic Press.

Popper, K. (1965). Prediction and prophecy in the social sciences. In *Conjectures and refutations: The growth of scientific knowledge* (pp. 336–346). New York, NY: Routledge.

Romanos, G. (1994). Reflexive predictions. In M. Martin & L. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 145–156). Cambridge: MIT Press. (Reprinted from *Philosophy of Science, 40,* 1973.)

Developmental Psychology

This entry looks into the social science of developmental psychology, traces the changes through which it has gone, focuses on its various epistemic aspects, such as research design or the theoretical issues associated with the nature/nurture debate, and goes on to introduce recent developments known as *neuroconstructivism* along with other new directions in the field.

All living organisms change over time. Developmental psychology is the scientific study of change in the perceptual, motor, social, emotional, linguistic, and cognitive capacities of humans. The field used to be mainly restricted to developmental change in the early years, from infancy to childhood; more recently, it has expanded to embrace atypical development (including genetic disorders such as Down syndrome, Williams syndrome, Autism Spectrum Disorder) and development throughout the entire life span (including the prenatal period and old age).

Human development is particularly slow compared with that of other species, and many scientists believe that this is an evolutionary advantage as it allows the human brain to adapt to environmental stimuli, to fine-tune learning, and to become progressively specialized over a lengthy developmental period. From a theoretical perspective, developmental psychology addresses numerous questions, two of which are crucial to understanding change over developmental time: (1) Are developmental changes gradual, or do they emerge rapidly with spurts of new forms of thought and behavior? (2) How do gene expression, brain function, cognitive processes, behavior, and environmental factors interact? The first question addresses the issue of continuity, that is, whether changes occur through progressive accumulation of knowledge or by developmental stages and through critical learning periods. The second question addresses the issue of the weighting of the

contribution of different factors to developmental change.

Developmental Research Design

Research in developmental psychology uses a wide range of different experimental and observational methods and covers both prenatal and postnatal participants. Modern neuroimaging techniques, such as functional magnetic resonance imaging (fMRI) or event-related potentials (ERP), allow researchers to study brain activity associated with a specific cognitive task in vivo. The use of powerful statistical procedures also enables developmental scientists to draw predictions and build models, or trajectories, of both typical and atypical development.

An ideal way to investigate development would be through longitudinal studies, in which researchers study the same individuals throughout a certain period of time. Due to time and cost constraints, cross-sectional designs are often preferred. This allows scientists to test people from different groups who share a similar history. The study of their families and environment also adds crucial information. Increasingly, a focus on individual differences rather than group differences is coming to the fore in developmental research.

Theoretical Issues: The Nature/Nurture Debate

One of the major issues of concern to developmental scientists is the way in which biological predispositions interact with the social environment. What is the relative contribution of nature and nurture to development? Two opposing theoretical perspectives address this relationship: the *biological maturation* and the *environmental learning* frameworks. Theorists of the biological maturation framework share the view that developmental changes are predominantly *endogenous*; that is, they arise from inside the organism in a sequence of genetically predetermined phases. Theories within the environmental learning framework see developmental changes predominantly as the product of *exogenous* causes: that is, they are shaped by environmental factors.

Interactionist Frameworks: Piaget and Vygotsky

The Swiss epistemologist Jean Piaget went beyond the long-standing nature versus nurture debate, addressing the issue of the dynamic of interaction between genes and the environment throughout development. He proposed the *constructivist* theory, according to which children construct sequentially higher levels of knowledge by actively engaging with their environment.

The Russian psychologist Lev Vygotsky also focused on interaction and the dynamic roles of nature and nurture. However, he introduced another factor: culture. In the cultural context framework, the interaction between genes and the environment is mediated by the knowledge transmitted from one generation to another through language, traditions, values, and beliefs.

Beyond Piaget

Piaget's theory of cognitive development was heavily criticized by Neo-Piagetians, in particular for not explaining individual differences in speed of processing and working memory, which regulate the child's ability to solve problems at different levels of complexity. Neo-Piagetian theories also integrate Vygotsky's social-cognitive principles to account for cultural and socioeconomic differences.

Neuroconstructivism

More recent approaches, such as *neuroconstructivism,* incorporate changes in brain circuitry over developmental time. Developmental scientists embracing this framework emphasize the importance of a multidisciplinary approach exploring developmental changes in gene expression, brain, cognition, and environment.

Developmental Change at the Genetic Level

Many studies attempt to map specific genes to specific behaviors. However, gene expression (where in the brain, and how much, protein is expressed) changes over time as a function of environmental influences. These epigenetic changes are becoming increasingly important for understanding development.

Developmental Change at the Neural Level

The structure and function of the human brain change significantly after birth. This is known as *neuroplasticity.* Using modern neuroimaging techniques, studies have shown that brain structure changes as a function of experience, for example, in learning a second language or developing specific skills even in adulthood.

Developmental Change at the Cognitive Level

To better understand cognitive development in normal conditions, modern research has also focused on neurodevelopmental disorders. Within the neuroconstructivist approach, cross-syndrome comparisons of cognitive abilities, such as language acquisition, number processing, and face processing, focus on tracing deficits at the cognitive level back to basic-level impairments in infancy, whose effects cascade over time on the developmental outcome.

Developmental Change at the Environmental Level

The environment is not static either. There are many factors that may affect development in both typical and atypical populations. Socioeconomic status and parental education are two of the most important factors, but modern developmental research also takes into account subtle differences in environmental factors, like the dynamics of interaction between children and caregivers.

New Directions

The nature/nurture debate continues to rage. Modern developmental psychology research has a renewed multidisciplinary focus on individual differences and also on important areas like bilingualism and multiculturalism. New brain imaging methodologies are constantly being developed and adapted to developmental populations, such as functional near-infrared spectroscopy, which provides better spatial resolution than ERP and better temporal resolution than fMRI. This noninvasive method measures neural activity in ways that are suitable for studying changes in brain development during early infancy.

The convergence of multiple methods as well as new experimental and statistical techniques should contribute to resolving the fascinating factors that make human development both similar to and different from that of other species.

Annette Karmiloff-Smith and Roberto Filippi

See also Cognitive Sciences; Evolutionary Psychology; Social Cognition; Social Neuroscience

Further Readings

Elman, J. L., Bates, E., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge: MIT Press.

Farran, E., & Karmiloff-Smith, A. (Eds.). (2012). *Neurodevelopmental disorders across the lifespan: A neuroconstructivist approach.* Oxford, England: Oxford University Press.

Johnson, M. H. (2005). *Developmental cognitive neuroscience.* Oxford, England: Blackwell.

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: Rethinking the study of developmental disorders. *Developmental Psychology, 45*(1), 56–63.

Karmiloff-Smith, A. (2010). Neuroimaging of the developing brain: Taking "developing" seriously. *Human Brain Mapping, 31*(6), 934–941.

Mora, S. (2007). *Cognitive development: Neo-Piagetian perspectives.* London, England: Psychology Press.

Piaget, J. (1952). *The origins of intelligence in children.* New York, NY: W. W. Norton.

Vygotsky, L. S. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Dialectic, in the History of Philosophy

Dialectic is a method or set of methods that is meant to provide a definition of a thing or a concept, which it achieves by reference to opposites in a discursive, step-by-step process of reasoning, through a number of questions and answers. Dialectic first appeared in ancient Greek philosophy and was rethought in its meaning and in its use throughout the history of philosophy, often in critical response to other thinkers. This entry traces the history of the notion of dialectic, and its role in philosophy from ancient times to our own.

Early History

The term *dialectic* (διαλεκτικής) comes from the Greek word *dialegomai,* "to converse with," and is used by Plato throughout his dialogues in order to establish the essence of any item whatsoever. As a set of argumentative methods, dialectic itself comes out of the live oral debates practiced by the Sophists, and most notably by Socrates, intended to demonstrate what a thing is. In Plato's early dialogues, dialectic is often an imitation of the Socratic questioning or discourse, *Sōkratikos logos,* or *elenchus,* which became an established literary genre also used by Socrates's other disciples, including Aeschines and Antisthenes.

Such dialectic is *negative* in that it leads to an apparent dead end or *aporia* in an *elenchic* or refutative dialogue, which, through a number of questions and answers, demonstrates the untenability of the initial thesis or definition. Taken at its extreme, negative dialectic may seem to suggest that any thesis and its antithesis can be equally refuted.

This notion of negative dialectic gives rise to the skepticism of the Middle and New Academy under Arcesilaus and the Stoic Carneades. It is in this sense that one needs to take Aristotle's ascription of the invention of dialectic to the Eleatic Zeno, who argued for the logical impossibility of the existence of the many, since one *can* demonstrate that "many" should be both finite and infinite at the same time. Yet in his later dialogues, Plato also uses dialectic as a positive method for achieving the "unalloyed knowledge" of the being of the forms (in the *Republic*) or of the one and the many (in the *Sophist* and *Parmenides*). As such, dialectic is a "coping stone" of all particular sciences, including mathematical sciences. The later Neo-Platonic philosopher Plotinus radicalizes this understanding of dialectic. For him, dialectic uses ordered reasoning to overcome and abandon discursive thinking, thus achieving Plato's "good beyond being," which is the source and condition of being yet itself is not being.

Dialectic in Plato does not appear as one single method but embraces a number of related logical methods. An example of such a method is *diairesis* ("division"), which establishes the sought-after definition through a number of consequent subdivisions of a concept (e.g., art) into several parts (principally in the later dialogues: *Sophist* or *Statesman*). One can distinguish other dialectical methods (analysis and synthesis, collection, definition, induction, *mesotēs,* i.e., searching for a middle between the extremes of more and less, etc.). Every such method is then only *a* method, a logical tool for achieving the knowledge of the essence of a thing. Since there is no prescribed way to obtain such knowledge, dialectic can be considered an *art, dialektikē tekhnē* (as in Plato's dialogue *Phaedrus*). However, the later tradition takes dialectic as a constitutive, logical part of philosophy, along with practical and theoretical philosophy (as in Alcinous's *Didaskalikos*). This division corresponds to Xenocrates's threefold division of philosophy into logic, ethics, and physics.

As a critique of Plato, Aristotle contrasts dialectic with syllogistic, the proper philosophical method, which alone is capable of understanding being as such, *to ontōs on*. Aristotelian dialectic is based on a seemingly correct opinion and considers only accidental properties; the premises of a dialectical syllogism are only plausible. For Aristotle, dialectic has much in common with rhetoric, insofar as neither considers a particular subject matter but provides logical form and strategies for argumentation.

In the Middle Ages, this led to the inclusion of dialectic, together with grammar, in the *trivium* (the tripartite core of standard medieval liberal studies: grammar, logic, rhetoric) and to the rethinking of dialectic as formal logic, for example, in Abelard's *Dialectica*.

Modern History

Early modern philosophy assimilates the logical subject that has predicates to subjectivity of the *ego* as the autonomous center of theoretical and moral activity. Such subjectivity should be equipped with a universal method of cognition that is seen in a logical method based on exhaustive enumerations arranged in a precise order. The ideal of the method is largely drawn from dialectic and developed in the work of Raimundus Lullus, Rudolph Agricola, and Petrus Ramus, and later in the work of René Descartes and G. W. Leibniz.

In the work of Immanuel Kant, *dialectic* becomes an integral part of his system of theoretical philosophy, but the term is employed with a significantly altered—negative in its connotation—meaning. The Transcendental Dialectic is a critique of reason that attempts in vain to go beyond the limits of possible experience, where reason aspires to achieve complete and absolute knowledge by unduly extending the use of the categories of understanding in an extra-empirical way, which leads to a transcendental illusion. After Kant, the search for a universal method for philosophy and science led German idealism to adopt dialectic as a method allowing for movement from one category to another within the edifice of a systematic philosophy. J. W. Fichte applies dialectic to producing categories out of the "I" and F. W. J. Schelling applies it to understanding nature.

In the work of G. W. F. Hegel, dialectic is given pride of place for becoming the method of building a grand synthesis of all logical, natural, and social phenomena within a philosophical system. Hegel's dialectical method describes the passage from one philosophical determination to another, from logic (being–essence–concept) to the philosophy

of nature to the philosophy of spirit (subjective spirit–objective spirit of institutions–absolute spirit of art, religion, and philosophy). Such development repeatedly occurs in the triadic structure of thesis–antithesis–synthesis (not his own terminology), where the synthetic member is sublated (in Hegel's terminology)—in other words both overcoming and preserving the content of the two previous opposite terms. Central to Hegelian dialectic are (a) postulating the dialectical motion by way of opposites not allowing for mediation, which creates a productive tension for the passage from one term to another, and (b) the distinction between reason (*Vernunft*) and understanding (*Verstand*), where speculative reason is a completed totality and alone is capable of grasping the opposites in their unity, inaccessible to understanding. Both principles are Platonic in origin. Before Hegel, these principles had been used by Nicholas of Cusa, who also accepted the coincidence of opposites in actual infinity or the "absolute maximum," which, due to the suspension of the principle of noncontradiction, can only be grasped nondiscursively by reason and is not accessible to discursive understanding.

In subsequent philosophy and social theory, dialectic often comes as a critical appropriation and rethinking of Hegelian dialectic. Thus, Friedrich Schleiermacher, who originally had construed dialectic as a theory of scientific construction that makes knowledge into an "absolute science," later turned to the Platonic dialogical understanding of dialectic as the art of conversation in the realm of pure thinking. Karl Marx applies Hegel's dialectic to the materialistic understanding of history, where opposites are taken as antagonistic classes whose conflict results in social, economic, and political progress.

Recent History

In his *Negative Dialectics,* Theodor Adorno speaks against the Hegelian totalization, stressing nonidentity and difference that cannot be dialectically overcome and subsumed under the universal. Similarly, in *Truth and Method,* Hans-Georg Gadamer rejects dialectic as a logical system-building method and instead accepts dialectic as the art of leading a genuine conversation and of questioning ever further, thereby returning to the Socratic conversational origin of the Platonic dialectic.

Dmitri Nikulin

See also Dialectic, in the Social Sciences; Frankfurt School and Critical Social Theory; Idealism

Further Readings

Adorno, T. W. (1973). *Negative dialectics.* London, England: Routledge. (Original work published 1966 in German as *Negative Dialektik*)

Alcinous. (1993). *The handbook of Platonism* (J. Dillon, Trans.). Oxford, England: Clarendon Press.

Bennett, J. (1974). *Kant's dialectic.* Cambridge, England: Cambridge University Press.

Gadamer, H.-G. (2006). *Truth and method.* London, England: Continuum Press. (Original work published 1960)

Kaldis, B. (2008). The question of Platonic division and modern epistemology. In J. Dillon & M.-E. Zovko (Eds.), *Platonism and other forms of intelligence* (pp. 59–79). Berlin, Germany: Academie Verlag.

Nikulin, D. (2010). *Dialectic and dialogue.* Stanford, CA: Stanford University Press.

Plato. (2010). *Dialogues of Plato* (B. Jowett, Trans.). Cambridge, England: Cambridge University Press.

Rosen, M. (1982). *Hegel's dialectic and its criticism.* Cambridge, England: Cambridge University Press.

Rossi, P. (2000). *Logic and the art of memory.* Chicago, IL: University of Chicago Press.

Dialectic, in the Social Sciences

This entry introduces the notion of dialectic and its use in social science, tracing its history with special emphasis on Marx's thought and the Marxist tradition, where the notion loomed large, and concludes by mentioning contemporary developments.

Historical Antecedents

Dialectic is a term from ancient Greek philosophy, denoting a style of argument used by Socrates in the works of Plato (and further specified by him as a certain methodological procedure for discovering definitions, as demonstrated in some of his later dialogues, e.g., *The Statesman*). It refers to a search for truth that proceeds successively through statement and refutation. The method uses logic and rationality rather than entrenched opposition or persuasive rhetoric. It is also described by Plato as a highly

abstract mode of thought accessible only to trained philosophers, because only they can release their minds from the concerns of the phenomenal world and thus know a realm of timeless truth.

This general method was subsequently formalized as positing hypotheses or theses (positive statements), antitheses (negating or contradicting statements), and—through the identification of error or inconsistency—syntheses (new positive statements arising from this process). The concept was revived in 18th-century German philosophical idealism by J. G. Fichte and G. W. F. Hegel. Hegel's own formulation of dialectic purported to grasp opposites in their unity or the positive in the negative. His particular contribution was to claim that in the movement of successive contradictions nothing was ever fully negated but rather was sublated, that is, retained and transcended (*Aufhebung*). Neither used the now famous but overly simplistic triadic formula, thesis–antithesis–synthesis, which originated in contemporary commentary. This triad has been erroneously attributed to Karl Marx and Friedrich Engels, who made the most famous use of dialectic or dialectics in modern times.

Marx's Critique of Hegel's Dialectic

The Hegelian dialectic was a *political* matter for Marx, because the interpretation of Hegel's philosophical legacy was itself a political issue in the German states of the 1830s and 1840s. Hegelian philosophy posited a historically evolving unity between the thought-world of concepts and the world of material experience, rather than a unity existing merely in the perceptions of a conscious individual. By incorporating history, social life, and an overtly political goal of human freedom into his thought, Hegel politicized philosophy.

In Marx's early years, overt political activity was not permitted in what were then aristocratic, monarchical, and highly authoritarian systems. Academic versions of political controversy were somewhat tolerated but were also manipulated for political ends. From 1842, Marx's engagement with Hegel was thus not that of an academic philosopher but rather a political engagement with philosophical issues as a matter of democratic and class-conscious radicalism. Marx's substantial quarrel with Hegel and Hegelians was to do with their politically debilitating view that dialectical motion in thought was somehow the origin of all motion in the world, human and otherwise. Marx referred to this disparagingly as a demiurge.

Marx rarely mentioned dialectic after 1844, when German Hegelians faded from his political horizon. Subsequently, he produced scathing critiques of other would-be Hegelians, such as P. J. Proudhon, author of a dialectical system of economic contradictions. Marx's major writings after that either do not mention dialectic at all or refer to it only briefly.

The Marxist Tradition

After 1859, Marx's friend and occasional collaborator Friedrich Engels undertook a lengthy series of popularizations and commentaries on his work. In these writings, dialectic was claimed to be a central concept in Marx's unique method, which Engels ultimately claimed was equal to, and a synthesis of, the methods of Hegel and Charles Darwin. Engels's clearest definition emerged in a manuscript work, published in Soviet Russia in 1925, in which he formulated three *materialist laws,* purporting to explain nature, history, and thought:

- Transformation of quantity into quality
- Interpenetration of opposites
- Negation of the negation

In that light Marx's occasional comments many years earlier have assumed an importance within this particular interpretive framework.

Marx's later comments on dialectic are in correspondence or in response to reviewers. In one of his most famous remarks, Marx simply said that his dialectic had a rational and political form, including in its comprehension and affirmation of existing conditions a critical recognition of their inevitable negation. In connection with his critical work on political economy, he commented that he had found Hegel's *Logic* very useful in analyzing and presenting conceptual relationships. However, he seldom identified as a materialist, having criticized traditional materialism for its reduction of human activities to timeless and history-less material factors. His new materialism, briefly identified as such, was founded instead on human social activities as they have evolved historically. He focused particularly on developments in industrial production and in corresponding social structures and forms of thought.

From selective interpretations of Marx's and Engels's works, the 20th-century philosophy of *dialectical materialism* evolved within communist and Soviet circles, but this was never accepted in the Western academy. Dissident Marxists influenced by Hegel, such as those in the Frankfurt School, promoted a nontriadic and nonmaterialist form of political criticism and social science in the 1920s and 1930s, using dialectic as a metaphor for their historicism and antipositivism.

Revivals and Developments

From the 1970s, there have been noncommunist revivals of dialectical thinking as a philosophy. These works link the polyvalent interrelatedness of concepts with the critical understanding of social phenomena, particularly capitalism. Dialectical approaches to society and politics have been somewhat overtaken since the 1990s by the "linguistic turn," in which Hegel and other German idealists and phenomenologists represent a major influence. However, this successor tradition also owes a considerable debt to the linguistic philosophy of J. L. Austin and the "language-games" of the later Ludwig Wittgenstein, so it cannot be considered dialectical in any straightforward way.

Terrell Carver

See also Dialectic, in the History of Philosophy; Frankfurt School and Critical Social Theory; Marxism and Social/Historical Explanation; Philosophy of History

Further Readings

Bhaskar, R. (2008). *Dialectic: The pulse of freedom*. New York, NY: Routledge.

Carver, T. (1998). *The postmodern Marx*. Manchester, England: Manchester University Press.

Jameson, F. (2010). *Valences of the dialectic*. London, England: Verso.

Levine, N. (1984). *Dialogue within the dialectic*. New York, NY: HarperCollins.

Marx, K. (1975). Critique of the Hegelian dialectic and philosophy as a whole. In K. Marx & F. Engels, *Collected works* (Vol. 3). London, England: Lawrence & Wishart.

Mészáros, I. (2011). *Social structure and forms of consciousness: Vol. 2. The dialectic of structure and history*. New York, NY: Monthly Review.

Ollman, B. (2003). *Dance of the dialectic: Steps in Marx's method*. Urbana-Champaign: University of Illinois Press.

Dialogical Logic

Dialogical logic (DL) is a novel field that aims at recovering the philosophical and technical links between argumentation, logic (logic as *Agon*), and epistemology, via the development of a pragmatist semantics. This semantics, inspired by both Ludwig Wittgenstein's *language-games* and game-theoretic concepts, provides the basis for the notion of formal strategy by means of which inference is understood dynamically—that is, as a kind of a rational interaction of agents.

The link between argumentation and logic, particularly in the context of the social sciences, may seem self-evident. Moreover, in the writings of modern thinkers, argumentation and argumentation practice have maintained the importance they gained in the time of the ancient Sophists. However, after the decline of formal axiomatics, the aforementioned tie with logic has declined.

This entry presents an overview of the new field of DL and its implications for the social sciences.

Background

In the late 1950s, Paul Lorenzen was the first to introduce a game-based theory of meaning for intuitionistic and classical logic, and it was further developed by Kuno Lorenz—somewhat resembling Socratic dialogues, Arabic theories of debate, and the medieval theory of *Obligationes*. At almost the same time as Lorenzen, Jaakko Hintikka developed a model-theoretical approach known in the literature as *GTS* (Game-Theoretical Semantics). Since then, a number of different game semantics have been studied in logic. Shahid Rahman and collaborators developed what they called *Dialogic* into a general framework for the study of logics such as connexive logic, epistemic logic, free logic, IF logic, linear logic, logic of belief revision, modal logic, nonmonotonic logic, paraconsistent logic, public announcement logic, and relevant logic.

This novel philosophical impulse experienced a parallel renewal in the fields of theoretical computer science and its interface with the social sciences, triggered by the work of Johan van Benthem and collaborators in Amsterdam, who explore thoroughly the interaction between logic and games. New results in linear logic by J.-Y. Girard in the interfaces between

mathematical game theory and logic, on the one hand, and argumentation theory and logic, on the other, resulted in the work of many others, including S. Abramsky, A. Blass, T. Coquand, D. Gabbay, M. Hyland, R. Jagadessan, G. Japaridze, E. Krabbe, H. Prakken, G. Sandu, T. Tulenheimo, D. Walton, and J. Woods, who placed game semantics in the center of a new approach in which logic is understood as a dynamic instrument of inference.

The point underlying dialogical-approach semantics is that those rules that fix meaning may be of more than one type and they determine the kind of reconstruction of an argumentative and/or linguistic practice that a certain sort of language-games called *dialogues* provide. The dialogical approach is not a logic but a *semantic rule–based framework* where different logics could be developed, combined, or compared.

More precisely, in a dialogue, two parties argue about a thesis respecting certain fixed rules. The player who states the thesis is called the Proponent (**P**); the rival is called the Opponent (**O**). Dialogues are designed in such a way that each of the plays ends after a finite number of moves, with one player winning while the other loses. Actions or moves in a dialogue are often understood as *utterances* or as *speech acts*. The point is that the rules of the dialogue do not operate on expressions or sentences isolated from the act of uttering them. The rules are divided into particle rules or rules for logical constants (*Partikelregeln*) and structural rules (*Rahmenregeln*). The structural rules determine the general course of a dialogue game, whereas the particle rules regulate those moves (or utterances) that are requests (to the moves of a rival) and those moves that are answers (to the requests).

Crucial for the dialogical approach and what distinguishes it from all other approaches are the following points (which will be elucidated later in the entry):

- The distinction between local (rules that determine how to challenge and defend a logical constant) and global meaning (general rules on how to play)
- The player independence of local meaning
- The distinction between the play level (winning of a play) and the strategic level (existence of a winning strategy)
- The notion of formal play

Dialogical Logic and Meaning

Local Meaning

Particle Rules: In dialogical logic, the particle rules are said to state the local semantics: What is at stake is only the request and the answer corresponding to the utterance of a given logical constant, rather than the whole context where the logical constant is embedded.

The following displays the particle rules, where X and Y stand for any of the players **O** or **P**:

Utterance:	**X:** $\alpha\vee\beta$ (X utters α *and/or* β//I////I//).
Challenge:	**Y:** ?-$\vee$ (Y requests X to utter one of both sides of the disjunction).
Defense:	**X:** α or **X:** β (X chooses).
Utterance:	**X:** $\alpha\wedge\beta$ (X utters α *and* β//I////I//).
Challenge:	**Y:** ?- $\wedge$L or **Y:** ?- $\wedge$R (Y requests X to utter the left or the right side of the conjunction).
Defense:	**X:** α or **X:** β (X chooses).
Utterance:	**X:** $\alpha\rightarrow\beta$ (X utters *If* α, *then* β).
Challenge:	**Y:** α (Y challenges the conditional by uttering α and requesting X to utter β).
Defense:	X: β.
Utterance:	**X:** $\sim\alpha$ (X utters *It is not the case that* α).
Challenge:	**Y:** α (Y challenges the negation by uttering himself α//I//).
Defense:	—(No defense is available; however, during the game, counterattacks on — might be possible).
Utterance:	**X:** $\forall x\alpha$ (X utters *For every x it is the case that* α).
Challenge:	**Y:** ?- $\forall$x/k (Y challenges the quantifier by choosing k and requesting X to utter αk//I//).
Defense:	**X:** α [x/k].
Utterance:	**X:** $\exists x\alpha$ (X utters *For some x it is the case that* α).
Challenge:	**Y:** ?- $\exists$ (Y requests X to choose one k for which α is the case).
Defense:	**X:** α [x/k].

Local meaning distinguishes between the following types of actions:

a. Choice of declarative utterances (=: disjunction and conjunction)

b. Choice of interrogative utterances involving individual constants (=: quantifiers)

c. Switch of the roles of defender and challenger (=: conditional and negation)

Player Independence: The particle rules are symmetric in the sense that they are player independent. If they were not, the logical constant would mean something for Player X and something else for Player Y. Take the case of chess as analogy: The rules that determine how, say, the knight moves, must hold for both players, the one playing with the white pieces as well as the one playing with the black pieces.

Global Meaning

Structural Rules

(SR 0) (Starting Rule): The initial formula is uttered by **P**. It provides the topic of the argumentation. Moves are alternately uttered by **P** and **O**. Each move that follows the initial formula is either a request or an answer.

(SR 1) (No-Delaying-Tactics Rule): After the move that sets the thesis players **O** and **P**, each chooses a natural number *n* and *m,* respectively (termed their *repetition ranks*). Thereafter, the players move alternately, each move being a request or an answer.

In the course of the dialogue, **O** (**P**) may attack or defend any single (token of an) utterance at most *n* (or *m*) times.

(SR 2) (Formal Rule): **P** may not utter an atomic formula unless **O** uttered it first. Atomic formulae cannot be challenged.

This rule allows us to formulate a notion of validity that does not amount to true in any model but is true *independently* of any model.

The dialogical framework is flexible enough to define what is known as *material dialogues:*

*(SR *2) (Rule for Material Dialogues):* Only atomic formulae standing for true propositions may be uttered. Atomic formulae standing for false propositions cannot be uttered.

(SR 3) (Winning Rule): X wins if it is Y's turn, but he cannot move (either challenge or defend).

Global Meaning

These rules determine the meaning of a formula where a particle occurs as a main operator in every possible play.

(SR 4i) (Intuitionist Rule): In any move, each player may challenge a (complex) formula uttered by his or her partner, or he or she may defend himself or herself against the last challenge that has not yet been defended.

(SR 4c) (Classical Rule): In any move, each player may challenge a (complex) formula uttered by his or her partner, or he or she may defend herself or himself against any challenge (including those challenges that have already been defended once).

Notice that the dialogical framework offers a fine-grained answer to the question: Are intuitionist and classical negation the same negations? Namely, the particle rules are the same, but it is the global meaning that changes.

In the dialogical approach, *validity* is defined via the notion of *winning strategy,* where winning strategy for X means that for any choice of moves by Y, X has at least one possible move at his or her disposal such that he or she (X) wins:

Validity (definition): A formula is "valid" in a certain dialogical system if **P** has a formal winning strategy for this formula.

Thus,

A is classically valid if there is a winning strategy for **P** in the formal dialogue Dc(*A*).

A is intuitionistically valid if there is a winning strategy for **P** in the formal dialogue Dint (*A*).

The main point of the dialogical approach is that dialogues are built from the bottom up, from local semantics (rules on how to challenge and defend a logical constant) to global semantics (rules on how to play) and from global semantics to validity via the notion of strategy (rules on how to win). This establishes the priority of the play level over the winning-strategy level. The levels are to be thought of as defining an order. From the dialogical point of view, to set the meaning of the logical constants via validity is like trying to define the (meaning) moves of the king in the game of chess by the strategic rules of how to win a play. Within the dialogical approach, the more basic step of meaning at the play level is the setting of player-independent particle rules.

Conclusion

One of the main objectives of DL is to motivate—overcoming the boundaries separating analytic

and continental conceptions of philosophy—a cross-fertilization between both computer-assisted approaches (including artificial intelligence) and formal approaches to the logic of argumentation, on the one hand, and informal approaches to reasoning, on the other. Furthermore, the dynamic approach to inference provided by DL should offer a theoretical framework that forges links between the humanities and social sciences in the central areas of concern to public life: informed debate, lucid decision making, and action based on reflection.

Shahid Rahman

See also Argumentation; Deduction; Inferentialism; Judgment Aggregation and the Discursive Dilemma; Language, Philosophy of; Language-Games and Forms of Life; Semantics and Pragmatics; Speech Acts

Further Readings

Blass, A. (1992). A game semantics for linear logic. *Annals of Pure and Applied Logic, 56,* 183–220, 1002.

Clerbout, N., Keiff, L., & Rahman, S. (2010). Dialogues and natural deduction. In G. Primiero & S. Rahman (Eds.), *Acts of knowledge, history, philosophy, logic* (chap. 4). London, England: College Publications.

Felscher, W. (1985). Dialogues as a foundation for intuitionistic logic. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (Vol. 3, pp. 341–372). Dordrecht, Netherlands: Kluwer.

Fiutek, V., Rückert, H., & Rahman, S. (2010). A dialogical semantics for Bonanno's system of belief revision. In P. Bour, M. Rebuschi, & L. Rollet (Eds.), *Constructions* (Vol. 14, pp. 315–334). London, England: College Publications.

Lorenz, K. (2011). *Logic, language and method: On polarities in human experience* (Gesammelte Aufsätze, 3 vols.). Berlin, Germany: De Gruyter.

Lorenzen, P., & Lorenz, K. (1978). *Dialogische logik*. Darmstadt, Germany: Wissenschaftliche Buchgesellschaft.

Rahman, S. (2009). A non-normal logic for a wonderful world and more. In J. van Benthem, G. Heinzmann, M. Rebuschi, & H. Visser (Eds.), *The age of alternative logics* (pp. 311–334). Dordrecht, Netherlands: Kluwer-Springer.

Rahman, S., Clerbout, N., & Keiff, L. (2009). Dialogues and natural deduction. In G. Primiero & S. Rahman (Eds.), *Acts of knowledge, history, philosophy, logic* (pp. 301–336). London, England: College Publications.

Rahman, S., & Keiff, L. (2004). On how to be a dialogician. In D. Vanderveken (Ed.), *Logic, thought and action* (pp. 359–408). Dordrecht, Netherlands: Kluwer.

Rahman, S., & Rückert, H. (Eds.). (2001). New perspectives in dialogical logic [Special issue]. *Synthese, 127.*

Rahman, S., & Tulenheimo, T. (2009). From games to dialogues and back: Towards a general frame for validity. In O. Majer, A.-V. Pietarinen, & T. Tulenheimo (Eds.), *Games: Unifying logic, language and philosophy* (pp. 153–209). Dordrecht, Netherlands: Springer.

Rückert, H. (2001). Why dialogical logic? In H. Wansing (Ed.), *Essays on non-classical logic* (pp. 165–185). London, England: World Scientific.

Disagreement

Disagreement is of interest to philosophers because it may seem to undermine the possibility of attaining knowledge, or rational, confident opinions, on controversial matters. This entry surveys the main lines of approach to this epistemological problem. Disagreement is, of course, a central feature of social life.

Disagreement is to be expected when different people have different information or when there is a clear disparity in cognitive skills. But difficult issues arise when disagreement occurs between "epistemic peers": people who seem to possess (roughly) equal cognitive skills and who are familiar with (roughly) the same information and/or arguments relevant to the disputed issue.

The intractability of certain kinds of disagreements—for example, ones about what tastes good and perhaps those involving artistic or moral judgments—are sometimes taken to indicate that the relevant subject matters are not really matters of objective fact. But epistemological problems arise from peer disagreement on questions that would seem to be straightforwardly factual. Persistent widespread disagreements of this type arise in history, economics, psychology, and philosophy—even among experts in the relevant fields. Peer disagreements over factual questions also occur in everyday social interactions, as when people disagree over who will win an election or about what their shares of an equally divided restaurant bill come to.

The main question raised by these disagreements is a *normative* one: How should knowing about disagreements by one's peers affect what one believes? Or, alternatively, what is the rational response to learning about peer disagreement? (Related questions also apply to disagreement by one's epistemic

superiors or inferiors, but to make it simpler, most of the discussion focuses on peer disagreement.)

Some have argued that persistent disagreement among experts in fields such as philosophy indicates that whatever methods the experts are using cannot be very reliable (if they were, the experts would tend to converge on the true answer). They have argued that recognizing this unreliability should preclude a person from having confidence in answers to controversial questions in these fields—even when some particular answer seems to her to be strongly supported by the evidence and arguments.

A similar position has been taken with respect to the local disagreements that crop up in ordinary social interactions. If a person finds herself disagreeing with someone she has good reason to regard as an epistemic peer, it is argued that she should realize that one of them has made a mistake, and she should take the possibility that she made the mistake as being at least roughly as likely as the possibility that her peer made the mistake. If she reasons this way, it seems that she should lose confidence in her original belief.

Positions of this sort—which claim that one should often lose confidence in the face of peer disagreement—are sometimes termed "conciliatory." An opposing sort of position holds that one may often rationally retain confidence in one's beliefs, even in the face of peer disagreement; positions of this sort are sometimes called "steadfast." (These labels, of course, refer to opposite directions on a spectrum of possible positions.)

Advocates of steadfast responses to peer disagreement point to the reasons one might have for privileging one's own initial belief over that of one's peer. The most prominent strand of steadfast argument points out that one's initial belief might in fact be well supported by one's initial evidence and arguments; that is, one's peer might have made the mistake. In that case, it is argued, the original evidence and arguments should tilt the balance in favor of one's original belief, and one needn't lose much (or, in the radically steadfast view, any) confidence in one's original belief.

Advocates of conciliatory views argue that this sort of steadfast response begs the question: One cannot rationally rely on one's own initial reasoning—the very reasoning apparently called into question by the peer's disagreement—to support the conclusion that one's peer was the one who made the mistake. On the other hand, advocates of steadfast views argue that conciliatory reactions to disagreement amount to irrationally ignoring or throwing away one's original evidence. Some also worry that conciliatory views lead to an unacceptable degree of skepticism.

It may be a source of worry that inquiry might suffer if people in general lost confidence in their views on controversial topics; perhaps, scientific research goes best when different investigators are committed to different lines of research. But this is compatible with seeing the individual investigator's confidence as irrational; it could be that certain patterns of individual irrationality are conducive to collective scientific progress.

Another avenue of approach to disagreement involves asking whether inquirers (or communities) who follow conciliatory policies do better or worse than inquirers who follow steadfast policies, when it comes to forming accurate beliefs. Some have explored this using mathematical models and computer simulations. This approach often takes beliefs to come in degrees of confidence and takes the aims of inquiry to include having high confidence in truths and low confidence in falsehoods. Different approaches to modeling may involve different assumptions about inquirers and different ways of measuring accuracy. The results of these exercises in modeling are mixed, with no general policy appearing clearly the best.

A more traditionally philosophical approach to the question relates it to other questions about rationally accommodating evidence that one has made a cognitive error. The disagreement of others is one source of such evidence, but one may also get evidence that one is biased, overtired, oxygen deprived, or under the influence of judgment-distorting drugs or psychological manipulation. In reacting to these bits of evidence, the same sort of question arises: If a thinker has initially reasoned correctly, can her rational confidence in her conclusion be undermined by her getting evidence that he's made a cognitive mistake? Again, arguments turn on whether particular responses beg the question, or disregard the original evidence.

David Christensen

See also Argumentation; Dialogical Logic; Epistemology; Philosophy of Expertise; Promises and Agreements

Further Readings

Christensen, D. (2009). Disagreement as evidence: The epistemology of controversy. *Philosophy Compass, 4,* 756–767.

Feldman, R., & Warfield, T. A. (Eds.). (2010). *Disagreement*. New York, NY: Oxford University Press.
Lackey, J. (2010). Epistemology of disagreement. *Oxford Bibliographies Online: Philosophy*. doi:10.1093/obo/9780195396577–0137

Disciplinarity

Disciplinarity may be taken to denote the presence of distinct and well-circumscribed modes of knowledge systematically organized into authoritative bodies of discrete specializations representing different "sciences" or "disciplines" cognitively and institutionally separated from each other. This entry presents aspects of disciplinarity, or lack thereof, dividing "disciplines" within the social sciences into different kinds along that line, and charts the possible historical and political causes behind the genesis of distinct disciplines in the social sciences.

Disciplinarity is less widely discussed than its nominal opposite, interdisciplinarity. For some social scientists, interdisciplinarity is a virtue, albeit one whose philosophical basis is difficult to specify, while disciplinarity is a vice, one whose meaning is clear. For others, adherence to disciplinary norms and methods is a mark of intellectual integrity, while interdisciplinarity is a byword for the inexact study of transient matters. Given the proliferation of (postmodernist) notions that celebrate approaches that favor the "multi-," "trans-," "inter-," "intra-," and so on, methodological routes of study privileging all sorts of "interface," to talk of or defend rigid disciplinarity—independent and self-bounded disciplines, especially in the social science—may sound outmoded to some critics. They claim that disciplinarity is an invention.

A disciplinary attitude is more likely in some social sciences than in others; in economics, political science, or some branches of psychology, a belief that confident generalizations—including causal laws—about social phenomena can be attained using intradisciplinary resources is still common. By contrast, sociology and anthropology have traditionally been more porous, less clearly specialized disciplines, not the least because they have often taken on the task of reflecting on the nature of the social sciences themselves. This distinction between degrees of disciplinarity is related to further distinctions between the degrees of rigidity with which social science disciplines define their object domain and methods, and the ease with which different object domains lend themselves or are claimed to lend themselves to firm intradisciplinary generalizations. Although internal disagreement is endemic to all social sciences, economists and political scientists have traditionally subscribed more readily to claims about the *laws* of the market or the perennial rules of political conflict (e.g., the German sociologist Robert Michel laid down what is known as his "iron law of oligarchy" in 1911) than sociologists or anthropologists have done to claims about the existence of social rules or historical laws.

One common claim in the philosophy of the social sciences is that if this is the case it is because the former disciplines (economics and political science) have an identity based upon an institutionalized body of theory and method, a stable corpus of empirical findings, a clearly bounded object domain, and a "core research program." The originator of the latter term, Imre Lakatos, believed that none of the social sciences were capable of generating one, and curiously enough, many such claims to disciplinary robustness in the social sciences have themselves depended on an appeal to models and modes of data processing taken from the natural sciences. Such models and procedures have occupied an intermediate and in some ways genuinely interdisciplinary zone, being available to a series of individual disciplines that have sought to import them with varying degrees of enthusiasm. Sociology and anthropology have remained the most internally divided social sciences in this respect, with structuralist, structural-functionalist, and systems theory sitting uneasily alongside hermeneutic or interpretive approaches; the sense of disciplinarity here is maintained by an uneasy agreement about the existence of "society" or "culture."

This raises the question of what holds disciplines together, how they maintain their sense of disciplinarity, and how they survive or adapt to historical change. The modern social science disciplines emerged during a particular historical period, adding to or replacing older disciplines, and there is no obvious reason why they should survive in their present form. "The light of cultural problems moves on," Max Weber said, and new object domains emerge, along with new disciplines to study them. Thus, it might be said that economics emerged with the growth of market capitalism; psychology with the growth of more interiorized understandings of the person, attendant on the rise of "bourgeois society"; and sociology with the making of the working class

and the rise of the "social question" in 19th-century Europe. On the other hand, there is much to the claim that those disciplines themselves emerged when they did as part of an internal development within a more strictly discursive field and that, moreover, they helped shape the emergence of those very objects of inquiry. In that case, disciplinarity—and the hard to specify relationship between modern disciplines and modern states—is a driving force of modern history as much as it is a response to it. The question of the precise weight to accord material and ideal factors in intellectual history is a perennial one, and it is illustrated when one considers the recently voiced claims that at the start of the 21st century market capitalism (object of economics) remains stronger than ever while "society" (object of sociology) is disappearing; these claims have been made not by economists but by sociologists themselves.

A final question is whether disciplinarity itself might one day disappear. There are indications that in some areas of the social sciences, disciplines defined by an institutionalized body of theory and method are not only being added to but also supplanted by new types of research activities. These activities both appeal to overtly non–natural-scientific models, notably those found in the humanities, and increasingly identify themselves according to substantive interests or concerns, defined tightly enough to suggest a research field but loosely enough to foster collaborative or interdisciplinary research efforts. University research centers, and sometimes whole departments named after such a substantive area—often going under the title of "X studies"—or the amalgamation of discipline-based departments into "schools," are an indicator of this trend.

The consequences for the social sciences of the increasing popularity of non–natural science methods are variable; the use of hermeneutic methods can appear incongruous in economics or political science or can be seen as an opportunity for innovation. By contrast, it may push disciplines such as sociology or anthropology toward the substantive agendas to be found in cultural studies or take them back to their early-20th-century classics, which still occupied an ill-defined zone between literature and science. It remains an open question whether this proves or disproves the maxim that a discipline that is afraid to forget its founders is doomed.

Charles Turner

See also Foucault's Thought; Interdisciplinarity; Lakatos, Methodology of Scientific Research Programs; Models in Social Science; Paradigms of Social Science; Postmodernism

Further Readings

Foucault, M. (1970). *The order of things*. London, England: Tavistock.

Lakatos, I. (1978). *Philosophical papers: Vol. 1. The methodology of scientific research programmes*. Cambridge, England: Cambridge University Press.

Lepenies, W. (1988). *Between literature and science: The rise of sociology*. Cambridge, England: Cambridge University Press.

Smith, R. (1997). *The Fontana history of the human sciences*. London, England: Fontana.

Wagner, P. (1999). An entirely new object of consciousness, of volition, of thought: The coming into being and (almost) passing away of "Society" as a scientific object. In L. Daston (Ed.), *Biographies of scientific objects* (pp. 132–157). Chicago, IL: University of Chicago Press.

Discourse Analysis

This entry presents an overview of the field of discourse analysis and all its ramifications, charts its multidisciplinary nature, shows how it has been intricately related to a number of social sciences, and shows its importance both for the humanities and the social sciences.

The Emergence of Discourse Studies Between 1964 and 1974

Discourse analysis, or discourse studies, is a cross-discipline that emerged between 1964 and 1974 in most of the disciplines of the humanities and the social sciences.

In anthropology, arguably the first discipline explicitly focusing on discourse (besides traditional literary studies and classical rhetoric), the ethnography of speaking, most notably represented in the work of Dell Hymes, examined "communicative events"—before the recent development of linguistic anthropology as such.

In linguistics, its main aim was to go beyond the sentence boundaries of structural and generative grammars, with the argument that real language use

takes place in the form of text and talk and not as isolated sentences.

The sociology of everyday life, and especially ethnomethodology, developed the very successful study of conversational interaction, soon also broadly represented in neighboring disciplines.

At the same time, in the early 1970s, cognitive psychology began its first experimental studies of discourse comprehension, while in the field of artificial intelligence, automatic processing of discourse began to be explored together with the representation of knowledge, crucially necessary in all production and comprehension of discourse.

Finally, in the 1980s, social psychology, especially in the United Kingdom, developed its own constructivist "discursive psychology." More recent are developments in communication studies, for instance, of media discourse analysis as well as the study of organizational discourse. Strangely, besides its traditional accounts of political rhetoric, political science has been among the few disciplines in the social sciences in which the explicit and systematic study of text and talk has not yet been introduced successfully. Parallel to these widespread empirical and concrete studies of text and talk in various disciplines, philosophy in the 1960s also knew various (obviously more conceptual) approaches to discourse, most notably represented in the work of Michel Foucault.

Neighboring Developments

It is interesting to observe that in the same period other inter- or cross-disciplines also emerged from within the humanities and the social sciences, often overlapping with discourse studies. Thus, *semiotics,* at first especially in literature, films, and the arts, was one of the developments in structural linguistics and related disciplines, later more broadly concerned with sign systems and semiosis, especially of nonverbal, multimodal messages. Whereas traditional, structural, and generative grammars and linguistics were limited to phonology, syntax, and (later) semantics, the *philosophy of language* proposed the study of *speech acts* and illocution, in what soon would be called *pragmatics*—closely related to discourse and conversation analysis, as was the case with the work of the philosophers John Austin, John Searle, and Paul Grice. The more empirical social and cognitive approaches to language use were the object of studies in the quickly developing interdisciplinary fields of sociolinguistics and psycholinguistics. Even formal (logical, etc.) grammars finally turned to the study of some properties of discourse, such as coherence and coreference, influenced, for example, by the work of the logician Richard Montague on grammar.

Methods and Dimensions of Discourse Studies

Today, the study of discourse has become a major area in virtually all disciplines of the humanities and the social sciences. Whereas earlier grammars hardly offered much more than an abstract and formal approach to some linguistic phenomena (e.g., language learning), multidisciplinary discourse studies offered a very broad and widely applied and applicable approach to the study of text and talk in many aspects of human and social life.

In the social sciences, the label "discourse analysis" is often understood as referring to a *method* of analysis, comparable with traditional *content analysis,* instead of to a (cross-)discipline with many different methods. This is also why the term *discourse studies* is currently preferred to refer to that discipline.

The study of discourse makes use of a variety of methods, many of these also common in the social sciences, such as (various types of) ethnography, participant observation, interviews, life histories, focus groups, document analysis, historiography, action and interaction analysis, laboratory and field experiments, and so on, represented elsewhere in this encyclopedia. It bears to be stressed what is usually forgotten in the application of these methods elsewhere in the social sciences—namely, that the means and products of these methods are mostly forms of text and talk, whose qualitative analysis again would require some kind of discourse analysis.

More typical of qualitative discourse studies is the systematic and explicit analysis of the structures or strategies of spoken or written discourse, for instance, at the following levels of description.

Grammar

Various types of structural, generative, and functional grammars are still dominant in the detailed description of the phonological, syntactic, and semantic structures of sentences. Discourse grammars have gone beyond the scope of the sentence and specifically

focus on local and global coherence, discourse topics (semantic macrostructures), the description of actors and agents, implications, presuppositions, and so on. This remains also the most systematic and explicit approach to discourse, as much of it can be based on quite sophisticated formal approaches. For the social sciences, a detailed *semantic analysis* of text or talk is one of the most interesting linguistic methods to be used in many forms of research, for instance, for the analysis of the ways polarized (Us vs. Them) ideologies are expressed and reproduced by discourse, for instance, by emphasizing Our Good Things and Their Bad Things (and mitigating Our Bad Things and Their Good Things).

Conversation Analysis

No doubt the most popular application of sociological discourse analysis has been the focus on everyday interaction in *conversation analysis*, initially interested in the study of turn-taking of everyday talk but later expanding to many other properties of situated conversation, also in organizational and institutional contexts, such as starting and ending conversations, topic change, storytelling, agreeing and disagreeing, and much more.

Narrative Analysis

Among the various genres and formats of discourse organization and function, narrative has probably been studied the most—besides conversation—beginning with traditional literary studies of narrative. Later, in sociolinguistics as well as in the other social sciences, the focus turned to everyday informal storytelling, for example, as a method of life history analysis.

Argumentation Analysis

Developing from classical rhetoric and dialectics, contemporary argumentation analysis also has found application in many other disciplines, for instance, in law (courtroom interaction) and science (scientific argument). Such analyses study the ways explicit and implicit arguments are formulated in support of a thesis or standpoint, according to the rules and moves of appropriate reasoning and argumentation, that is, avoiding the usual fallacies.

Genre Analysis

Discourse comes in many types or genres, basically defined in terms of the *context* of text and talk, consisting of *setting* (time, place), *participants* (and their *social and communicative identities, roles, and relations*), a *social action* being accomplished by the discourse (e.g., giving a lecture, consulting a doctor), and the *goals and knowledge* of the participants. Such genre analyses systematically study the linguistic and other formal and functional properties of the many kinds of text and talk used in the domains of politics (e.g., parliamentary debates), the mass media (e.g., news, editorials, or soap operas), education (e.g., textbooks and classroom interaction), the bureaucracy (e.g., forms, interaction with clients), the law (e.g., laws, interrogations, testimony), and science (e.g., scholarly articles, reviews), among many others.

Multimodal (Semiotic) Analysis

Discourse is not just verbal but, with the introduction of the computer and the Internet, increasingly multimodal, combining verbal structures with images, sound, and so on. Contemporary social semiotics is developing sophisticated theories for the systematic analysis of such multimodal messages.

Stylistics and Rhetoric

Traditionally associated with literary studies, the study of style and rhetoric also has developed more broadly across disciplinary boundaries, for instance, for the study of the linguistic variation of the syntax and lexicon of text and talk and for the classical ways discourse can be made more persuasive with rhetorical structures ("figures of speech") such as rhyme, repetition, hyperbole, euphemism, metaphor, and metonymy, among many others, functioning to enhance or mitigate the meaning of discourse. The polarized (Us vs. Them) structures of ideological discourse typically use various forms of rhetorical "figures," such as hyperboles and euphemisms, to emphasize or mitigate assumed in-group and out-group properties.

Corpus Analysis

Computers and databases have enabled the storage of vast corpora of text and talk. These corpora allow quantitative and qualitative analyses of thousands of discourses and many millions of words, word combinations and collocations, keywords, syntactic structures, and so on. Especially for the study of large numbers of texts, such analyses have become indispensable in much contemporary discourse studies.

Conclusion

Since its striking emergence in the humanities and social sciences half a century ago, the cross-discipline of discourse studies has come a long way. There is virtually no form of human conduct, interaction, and communication that does not involve the use of language as it is manifested in a vast number of types of text or talk. Multidisciplinary theories and increasingly sophisticated methods of *quantitative* and especially *qualitative* discourse analysis have been developed. These have found widespread application in the study of a large number of social problems, such as the discursive nature of the acquisition and reproduction of ethnic prejudice and racism. More than any other development in the humanities and social sciences, the study of discourse has established multidisciplinary relationships among these disciplines. Discourse is at the same time a form of language use, communication, interaction, social cognition, and power—and hence the interface between micro- and macrostructures of society.

Teun A. van Dijk

See also Communication Studies; Ethnomethodology; Foucault's Thought; Holism, in the Philosophy of Language; Language and Society; Semantics and Pragmatics; Sociolinguistics; Speech Acts

Further Readings

Blommaert, J. (2004). *Discourse: A critical introduction.* New York, NY: Cambridge University Press.

Jaworski, A., & Coupland, N. (1999). *The discourse reader.* London, England: Routledge.

Renkema, J. (2004). *Introduction to discourse studies* (2nd ed.). Philadelphia, PA: John Benjamins.

Schiffrin, D., Tannen, D., & Hamilton, H. E. (Eds.). (2001). *The handbook of discourse analysis.* Malden, MA: Blackwell.

Stivers, T., & Sidnell, J. (Eds.). (2011). *Handbook of conversation analysis.* Oxford, England: Wiley-Blackwell.

Ten Have, P. (2007). *Doing conversation analysis: A practical guide* (2nd ed.). Thousand Oaks, CA: Sage.

Titscher, S., Meyer, M., Wodak, R., & Vetter, E. (2000). *Methods of text and discourse analysis.* Thousand Oaks, CA: Sage.

van Dijk, T. A. (Ed.). (2007). *Discourse studies* (5 vols.; Sage Benchmark Series). Thousand Oaks, CA: Sage.

van Dijk, T. A. (Ed.). (2011). *Discourse studies: A multidiscplinary introduction* (2nd ed.). Thousand Oaks, CA: Sage.

Wodak, R., & Meyer, M. (Eds.). (2001). *Methods of critical discourse analysis.* Thousand Oaks, CA: Sage.

Wooffitt, R. (2005). *Conversation analysis and discourse analysis: A comparative and critical introduction.* Thousand Oaks, CA: Sage.

Distributed Cognition and Extended-Mind Theory

It may seem natural to think of the mind as a stream of conscious experience occurring squarely behind the eyes or perhaps as some single, persisting subject of these experiences, hovering in the center of the skull. The past 100 years of scientific thinking about the mind have challenged this view in a variety of ways. The most recent challenge, and the most striking to date, rests partly on the distributed nature of cognition—the fact that intelligent behavior emerges from the interaction of a variety of elements, some of which may be spatially removed from the locus of behavior. From such distributed models of cognition, many authors have inferred the extended-mind thesis, the claim that the mind itself spreads into the world beyond the boundary of the human organism.

Distributed cognitive models have had a direct impact on, and to some extent have been inspired by, research in the social sciences. Studies of insect behavior, for instance, form a bridge between the interests of cognitive scientists and matters to do with group-level behavior: Large numbers of social insects, each of which "mindlessly" follows simple information-processing rules such as "Drop my ball of mud where the pheromonal concentration is the strongest," design elaborate nests. This illustrates both how intelligent-looking results can arise from a distributed—and one might think fairly unintelligent—process and also how their emergence might be social in nature: Environmental conditions induce various subpopulations to play different roles in the life of the insect colony. Some robustly cognitive, human social processes also seem amenable to distributed theorizing: Contemporary scientific results, in particle physics, for example, often involve the contribution of hundreds, or even thousands, of individuals and instruments; here, each individual exercises a rich set of her own cognitive resources while playing a role in a much larger, highly structured enterprise. An intermediate case might be the modeling of traffic patterns: Individual humans can reason in flexible

and complex ways about driving and routes of travel, but constrained by the presence of other automobiles and the surrounding infrastructure, drivers' contributions to traffic flow, and the resulting traffic patterns, have much in common with the large-scale behavioral patterns of social insects.

The remainder of this entry consists of three sections. The first describes distributed cognitive modeling and the extended-mind thesis in more detail. The next section reviews critical reactions to the extended-mind thesis. The final section briefly discusses fruitful areas of ongoing research on distributed cognition and the extended mind.

From Distributed Cognition to the Extended Mind

In the early 20th century, Sigmund Freud proposed that subconscious mental states sometimes drive human behavior, thereby taking a significant step away from the commonsense view of the mind bruited at the outset. Given that the Freudian subconscious deals primarily in emotions and desires, however, Freudian innovations might seem to have little to do with intelligence or rationality per se. In contrast, the cognitivist revolution and its computer model of the mind marked a wholesale departure from the Cartesian view, at least with regard to the role of introspection in the life of reason. The new cognitive science set out to model the mechanisms the operation of which accounts for central aspects of human intelligence—memory, reading, means–end reasoning, perception, speech processing, and so on—with no particular emphasis on consciousness. The overriding goal was to model behavioral data or produce simulations of human-grade performance; it mattered not whether the details of such models jibe with the revelations of introspection. From the standpoint of the new cognitive science, all that matters is that a series of mathematically defined states, causally connected by precisely defined operations, produces intelligent behavior.

The demotion of consciousness together with the focus on mechanisms paved the way for distributed accounts of cognition. If one need not ground one's models in a conscious, first-person perspective, one is free to look for cognitive mechanisms wherever the causal contributors to human behavior might be. Think again of the historical context. The computer model of the mind enjoyed substantial empirical success, but much of this was in artificially limited domains or in the solution of narrowly circumscribed problems. Special-purpose systems can diagnose patients when given lists of symptoms, but such systems do not exhibit human-like intelligence, flexibly and smoothly navigating their way through the real world, in real time.

In response, a new breed of cognitive scientists began searching for a richer—or at least different—set of causal contributors to intelligent behavior—influences that might help produce human behavior without requiring that their contributions take the form of computer code. Much of this research focused on the entire, interactive system of brain, body, and world and, in doing so, cast the constraining influence of environmental structure in a new light. In addition to being a source of input, the environment appears to work together with bodily processes to guide intelligent behavior—both literally, as in navigation, and metaphorically, as when it limits a decision maker's options. The child does not learn to walk via the maturation-based appearance of an innate, internally stored motor program, for example; rather, walking emerges as the joint activity of the child's musculoskeletal system, the solid floor, and gravitational fields, all interacting as physical contributors to a stable cycle of movement through space. The visual system does not explicitly encode the proposition that objects in the environment are likely to be relatively rigid. Instead, the visual system computes the layout of the environment in a way that works effectively only in environments in which most of the objects are rigid. The rigidity assumption is left implicit, and thus, when the visual system functions properly, it does so by pooling its own contributions together with those of an environment that contains rigid objects. Philosophers took notice of, and sometimes contributed to, this developing trend in cognitive science, arguing, for example, that from an evolutionary perspective, we should expect environment-dependent cognitive processing to be widespread.

Into this intellectual foment, Andy Clark and David Chalmers cast "The Extended Mind" and, in doing so, crystallized the vanguard's vision. This vision draws heavily on empirical work related to distributed cognition. For example, a central notion from the literature on distributed cognition is that of an epistemic action. Such actions do not bring the physical world any closer to a state in which the subject's goal has been achieved; instead, epistemic

actions divulge information about how the subject might best pursue her goal. The rearranging of Scrabble tiles on one's rack, for instance, does not change the state of the game board so that it is any closer to having a new, completed word on it; rearranging the letters adds nothing to the board itself. Rather, this epistemic action provides information to the player's cognitive mechanisms so that she might more readily identify a word among her set of letters. Such actions take place partly in the environment, and thus we can see how elements of the problem-solving process itself might be outside the organism. What matters with regard to cognitive-scientific explanation is the role these contributors play in the production of intelligent behavior, not whether the contributors happen to appear within the boundary of skin and skull. Given the plausible additional premise that at least part of one's mind is physically located where one's cognitive processes are located, empirical work on distributed cognitive processing supports the extended-mind thesis.

The preceding style of reasoning also applies to pedestrian mental states such as belief. Perhaps the contents of most of a typical individual's beliefs are neurally encoded. Nevertheless, such contents could be encoded externally—in a trusty notebook, for example—and if these external encodings were to interact with the organism in the right way, they would constitute part of the material basis of human mental states. So long as (a) the subject carries the notebook with her—the information in it reliably available—(b) she trusts what's written in the notebook, and (c) she can fluidly deploy the information stored there, the notebook does not differ relevantly from internal resources, at least with regard to its role in the production of intelligent behavior. Thus, mental states recognized by our traditional, commonsense understanding of ourselves—beliefs, memories, perceptions—can reside, at least partly, beyond the boundary of the human organism. Moreover, new cognition-enhancing technological devices appear almost daily, and thus we can expect this phenomenon to become ever more pervasive.

Critical Reaction

Sometimes proponents of the extended-mind thesis press a merely prejudice-removing agenda; they hope to neutralize readers' internalist biases by showing how cognition and mind *could* comprise environmental states or processes. Illustrative examples (such as the Scrabble example given above) play an indispensable role in this project, but these illustrations often lead double-lives, being meant also to suggest that extended cognition surrounds us in the actual world. Here is where the more substantial question lies: Do cognition and mind actually extend into the environment in deep and substantial ways, extensively enough to ground a revolution in cognitive science and a reconception of ourselves as persons?

Advocates for the extended view typically claim that (a) it constitutes the best interpretation of existing cognitive-scientific results, (b) it offers the most promising framework for future research in cognitive science (even if the existing results don't conclusively support the extended view over its rivals), or (c) it captures the deep nature of the mind—where the plausibility of (c) derives from the plausibility of (a) or (b). It is no surprise, then, that critics often attack the cognitive-scientific basis of the extended-mind thesis.

In this vein, some critics have charged that an extension-friendly cognitive science is bound to lack scientific unity. External objects and states, and the processes in which they participate, differ significantly from the internal ones heretofore fruitfully studied in cognitive science. The behaviorally relevant causal profile of neurally encoded memories, for instance, differs enormously from the causal profile of, say, graphite scratches in a notebook. It may well be useful to study the complex ways in which humans interact with and exploit the environment during problem solving, but the external processes so studied will almost certainly differ in substantive ways from internal ones—at least if attention is directed to real-life cases involving paradigmatic cognitive capacities. As a result, there will be no reason to treat the entire run of states and processes, internal and external, as cognitive—as being of essentially the same kind. This thought is closely related to another: Even when the organism enters into an ongoing causal relation with portions of the environment, that does not by itself extend cognition or mind into the environment. Reasoning to the contrary commits the so-called causal-constitution (or coupling-constitution) fallacy.

Thus, it is demanded that philosophers of cognitive science produce a mark of the cognitive or at the very least some cluster of central traits that tend to qualify a state or process as genuinely cognitive. By adverting to such a mark, proponents of the

extended view might explain what unifies internal and external processes during extended cognitive processing and also make clear the conditions under which an ongoing causal connection brings into existence a fully cognitive process.

Critics themselves have proposed various marks of the cognitive—or at least necessary conditions on something's being genuinely cognitive—to ground their arguments against the extended-mind hypothesis or to undercut arguments in support of the extended view. Mental representations, for example, have played a central theoretical role in cognitive science. Perhaps, then, for a process to be genuinely cognitive, it must consist of the manipulation of representational states—that is, states that picture, refer to, or describe parts of the world. Even if some external units are representational, their representational status derives from our interpretation of them, which is believed to show that external processes are not cognitive in their own right.

A further critical approach contrasts the extended view with others that fall under the general rubric of *situated cognition*. According to one of these, the embodied view, nonneural bodily states or processes constitute significant parts of human cognition. According to another one, the embedded view, human cognitive processing is highly interactive and exploits structure in the environment, even though the external materials don't become proper parts of cognition itself. The critic claims, though, that data thought to support the extended view should be taken to support the embodied or embedded view instead. Once we have established the utility of a persisting cognitive system located in the body and have explained how it interacts with the passing parade of external stimuli and materials, it appears gratuitous to say that when the integrated, persisting cognitive system—typically housed within the body—interacts with the external materials, a new, fully cognitive system comes into being.

For all that's been said, this critical approach might seem to be a washout: The relevant data cohere well enough with either an extended or an embedded–embodied view. But this would be to ignore the central theoretical role of the relatively integrated, persisting *system*. It is the core construct in cognitive-scientific modeling and, as such, seems to offer the most promising basis on which to distinguish merely causal contributors to intelligent behavior from those processes the location of which bears on the location of the mind. All forms of cognitive modeling—computationalist, connectionist, dynamicist, as well as the brute-biological—specify a set of integrated elements and operations governing their interaction; the functioning of this system explains intelligent behavior as the result of interaction between the system and whatever materials or stimuli it happens to come across. On this view, then, being a state of a part of the integrated, relatively persisting system serves as a necessary condition for something's being a genuinely cognitive state. This constitutes a criticism of the extended view only because, and to the extent that, for humans, the integrated, relatively persisting cognitive system appears within the organism.

The consideration of a final critical perspective will bring more clearly into focus the relation between cognition and mind. One might accept that cognition is distributed but think that our enduring and stable conception of the mind should play a winnowing role when we evaluate the implications of distributed models of cognition. It is arguably a central feature of our concept of the mind that it be the locus of control of intelligent behavior. Thus, even if we bracket the demand for a mark of the cognitive and even if the persisting cognitive system stands as merely one element of many in cognitive-scientific modeling, this system has privileged status: It serves as the locus of control of the various forms of behavior cognitive scientists hope to explain; therefore, where we find it, we find the mind. If the integrated, relatively persisting cognitive system appears within the human organism—as seems likely to be the case for most humans, most of the time—then so does the human mind.

Future Directions

Debate about distributed cognition and its philosophical implications remains lively, as does the development and testing of distributed models. Among others, the following areas of research—of particular note in connection with the social sciences—flourish.

Some authors have attempted to characterize mathematically the relevant structural properties of integrated intelligent systems. According to one suggestion, a wide range of networks—including the brain and well-functioning computer networks—exhibit a distinctive trade-off between the specialization of

their components and the sharing of information among those components.

Along similar lines, structural properties have been sought at the level of social groups. For example, memory researchers have found that people who work together to solve problems often take on specialized roles within the group (e.g., with regard to what sort of information they're responsible for remembering); the distribution of such roles facilitates some efforts but may hinder others. This sort of investigation might help us decide whether group minds appear among the extended minds. It also has the potential to contribute substantially to research in the social sciences—from the analysis of voting behavior to the study of the dynamics of online communities.

Last, it's worth emphasizing the contribution dynamical-systems–based approaches have made to the discussion of distributed cognition and the extended-mind thesis. Computer-based and neural-network–based metaphors have dominated research in cognitive science. Nevertheless, a significant minority of cognitive scientists instead view the mind as something more like a physical system in motion, one best described using differential equations that express dependence relations among various continuously interacting quantities. On this view, cognition occurs when the organism enters, even for a short time, into an interdependent relation with physical systems beyond the boundary of the organism—with each component in the system constraining the other in a relationship of ongoing mutual determination. Proponents of such dynamical modeling often claim that cognition and mind permeate the entire system: the organism and the external materials to which it is coupled. It remains to be seen whether the dynamical-systems approach can be used to model paradigmatically cognitive activities such as long-term planning and theoretical reasoning. Nevertheless, exploration of dynamical processes should help us better understand at least some cognition-related interaction between thinkers and their environments.

Robert D. Rupert

See also Classical Computationalism, Connectionism, and Computational Neuroscience; Cognitive Sciences; Collective Agents; Collective Intentionality; Consciousness; Embodied Cognition; Group Mind; Mind–Body Relation; Situated Cognition

Further Readings

Adams, F., & Aizawa, K. (2008). *The bounds of cognition.* Malden, MA: Blackwell.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension.* New York, NY: Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19.

Hurley, S. L. (1998). *Consciousness in action.* Cambridge, MA: Harvard University Press.

Hutchins, E. (1995). *Cognition in the wild.* Cambridge: MIT Press.

Menary, R. (Ed.). (2010). *The extended mind.* Cambridge: MIT Press.

Noë, A. (2004). *Action in perception.* Cambridge: MIT Press.

Ross, D., Spurrett, D., Kincaid, H., & Stephens, G. L. (Eds.). (2007). *Distributed cognition and the will: Individual volition and social context.* Cambridge: MIT Press.

Rowlands, M. (1999). *The body in mind: Understanding cognitive processes.* Cambridge, England: Cambridge University Press.

Rupert, R. D. (2009). *Cognitive systems and the extended mind.* Oxford, England: Oxford University Press.

Rupert, R. D. (2011). Cognitive systems and the supersized mind. *Philosophical Studies, 152,* 427–436.

Tollefsen, D. (2007). From extended mind to collective mind. *Cognitive Systems Research, 7,* 140–150.

Wheeler, M. (2005). *Reconstructing the cognitive world: The next step.* Cambridge: MIT Press.

Wilson, R. A. (2004). *Boundaries of the mind: The individual in the fragile sciences.* Cambridge, England: Cambridge University Press.

Duhem-Quine Thesis and the Social Sciences

The so-called Duhem-Quine thesis emerged in the context of Willard Van Orman Quine's challenge to logical positivism/empiricism, first published in his famous "Two Dogmas of Empiricism" in 1951. This challenging thesis is called the Duhem-Quine thesis, not the Quine thesis, because Quine in footnotes attributes it to Pierre Duhem (1861–1916), a French theorist of thermodynamics with an in-depth knowledge of the history of science, who popularized the idea behind the thesis in the memorable phrase "saving

the phenomena"; that is, any body of observable facts can be sustained by suitably adjusted rival explanations against possible counterexamples. Duhem distinguished the physical sciences from metaphysics. Metaphysics, not science, explains the observable world. Thus, contrary to scientific realism, which claims that science explains the empirical world, Duhem argues that physical theory is merely a system of mathematical equations that sums up experimental laws. Despite Quine, it should be noted that the consensus among Duhemian scholars is that Duhem's views should not be conflated with those of Quine. Nonetheless, the term *Duhem-Quine thesis* remains standard terminology for Quine's thesis.

Centrally, the Duhem-Quine thesis concerns the experimental testing of scientific hypotheses. The thesis challenges two widespread claims: (1) the Baconian view of crucial experiments and (2) Popperian falsifiability. Contrary to the Baconian view, there are no crucial experiments. The experimental confirmation of a crucial prediction, such as the experimental confirmation that light rays are deflected by a gravitational field—a result crucial to the acceptance of Einstein's general theory of relativity—does not constitute a telling proof of that theory. Second, contrary to what the naive Popperians believed, the experimental disconfirmation of the logical consequence of an individual hypothesis does not necessarily falsify that hypothesis. Both Baconians and Popperians assume that a hypothesis can be tested on its own. This is not so: Hypotheses do not face the tribunal of experience individually; rather, they face it as part of a complex network including auxiliary hypotheses and theoretical presuppositions regarding the working of instruments, and so on. Hence, any hypothesis can he held true, irrespective of what may happen experimentally, provided scientists make appropriate adjustments, however drastic, elsewhere in the system.

In this entry, the Duhem-Quine thesis is examined within the context of Quine's philosophy of science, in particular his holism, and its interpretation within the social sciences is addressed.

Quine's Holistic Empiricism

The Duhem-Quine thesis is a consequence of Quine's innovative, holist linguistic approach to empiricism, which differs from classical Humean empiricism (after David Hume, with his separation of matters of fact from relations of ideas and consequently the epistemic revisability of the former but not of the latter—underpinning the rigid distinction between the analytic/synthetic distinction) and its 20th-century reformulation in logical positivism. Contrary to Humeans, the first Quinean step shifts the epistemological focus away from ideas and other mental entities onto words. The second step is the shift from words to sentences. The basic unit of communication is not the word; rather, it is the sentence. For example, if one does not know how to use the word *doctor* in a range of sentences, then one does not properly understand that term. Third, sentences are not atomistic units. A sentence is embedded in a system of sentences. For instance, to understand the sentence "The doctor prescribed ampicillin," one has to grasp a network of sentences about medicine. The next two steps are Quine's rejection of the "two dogmas" of classical empiricism: the analytic/synthetic distinction and the doctrine of reductionism. The former divides all significant discourse into logical truths (analytic sentences or relations of ideas, in Hume's language) and observational-empirical truths (synthetic sentences or sentences reporting matters of fact). Reductionism requires that complex synthetic sentences be translatable into basic sentences about the final elements of experience.

Quine's holistic philosophy implies that the totality of human knowledge, from, say, geography or history to pure mathematics and quantum physics, is an intricate, humanly constructed network of sentences that impinges on experience only along its edges—that is, a web of beliefs. This vast system is so underdetermined by sensory experience that rational beings have a genuine choice as to what sentences to reevaluate in light of a single piece of contrary evidence (so revisability is no longer constrained by the Humean distinction—followed by the logical empiricists too—between revisable, contingent knowledge of matters of fact and unrevisable, necessary, analytic sentences expressing relations of ideas). Thus, the rational standard for evaluating basic changes to a system is not the realist standard of correspondence to reality (of synthetic statements expressing matters of fact); rather, the standard is a pragmatic one and, thus, can touch even the so-called core of what were erstwhile thought of as necessarily true analytic statements (expressing unrevisable relations of ideas).

Interpretations of the Duhem-Quine Thesis and the Social Sciences

In general, by challenging the hegemony of logical positivism from within, the Duhem-Quine thesis contributed to opening up the methodology of the social sciences to what Bruce Caldwell and others call *methodological pluralism*. However, its most direct influence perhaps is evident in one of the major social sciences, namely, economics. A central concern in the philosophy of economics is the issue of the testing of economic theory, especially what is called neoclassical or orthodox theory. Following the marginalist revolution, neoclassical theory has proved to be very robust throughout the 20th century and into the next, despite its critics. Defenders of neoclassical theory, such as the American philosopher of economics Daniel Hausman, argue that the robustness of neoclassical theory is justifiable on the grounds of the Duhem-Quine thesis: The core of neoclassical theory is privileged by making adjustments in noncore areas. This defensive reading of the Duhem-Quine thesis is challenged by others using the more nuanced reading of the Duhem-Quine thesis provided by the British philosopher of science Mary Hesse. In light of Quine's own commitment to an empiricism without the dogmas of logical positivism, in Hesse's nuanced empiricist reading, no scientific, theoretical core is privileged—that is, immune from rejection. The core of a scientific theory is rejected if the totality of evidence, ranging from the empirical end to the analytico-conceptual end of the spectrum of evidence, tells against it. While no one single piece of empirical evidence taken on its own is capable of bringing down the core, the totality of evidence has this capability: Otherwise, the theory is not an empirical science. In this nuanced reading, the core of neoclassical economics is not privileged. Clearly, the Duhem-Quine thesis, given the range of interpretations from the defensive reading to the nuanced empiricist reading, continues to present challenges to philosophers of social science in their efforts at understanding the role of experience in the scientific endeavor.

Thomas A. Boylan and Paschal O'Gorman

See also Analytic/Synthetic Distinction; Critical Realism in Economics; Empiricism; Falsifiability; Holism, in the Philosophy of Language; Logical Positivism/Logical Empiricism; Pragmatism; Realism and Anti-Realism in the Social Sciences

Further Readings

Boylan, T. A., & O'Gorman, P. F. (2003). Pragmatism in economic methodology: The Duhem-Quine thesis revisited. *Foundations of Science, 8,* 3–21.

Caldwell, B. J. (1982). *Beyond positivism: Economic methodology in the twentieth century.* London, England: Routledge.

Duhem, P. (1977). *The aim and structure of physical theory* (P. P. Wiener, Trans.). New York, NY: Atheneum. (Original work published 1906)

Hausman, D. (1992). *The inexact and separate science of economics.* Cambridge, England: Cambridge University Press.

Hesse, M. (1970). Quine and a new empiricism. In G. N. A. Vessey (Ed.), *Royal Institute of Philosophy lectures 3: Knowledge and necessity* (pp. 191–209). London, England: Macmillan.

Quine, W. V. O. (1961). Two dogmas of empiricism. In W. V. O. Quine (Ed.), *From a logical point of view* (pp. 20–46). New York, NY: Harper & Row. (Original work published 1951)

Durkheim's Philosophy of Social Science

Émile Durkheim's philosophy of social science defines the goals and methods of his sociology. He sought to explain *social facts* in terms of social causes, which were to be discovered through testing causal hypotheses against the facts. His sociology was an inspiration for a variety of social sciences, including the French ethnological tradition of Claude Lévi-Strauss, the British social anthropology of Alfred Radcliffe-Brown, and the American structural functionalist sociology of Talcott Parsons. Social scientists have interpreted Durkheim's methodology each in their own way to suit their particular intellectual agendas.

Rather than investigate all the ways in which Durkheim's thought has been used or abused, this entry explains his philosophy of social science on its own terms. It considers the relationships among his methodology, sociological realism, and model of explanation against the background provided by the methods he actually used in his empirical sociological works.

Background

Durkheim (1858–1917) was a humanistically educated philosopher who sought to transform sociology into an empirical science that could provide answers to many of the questions that have traditionally been the domain of philosophy. He and his collaborators helped establish sociology as an academic discipline, drawing on works in philosophy, history, comparative law, economics, criminology, political geography, and ethnography. Although he had little first-hand knowledge of the natural sciences, he took them as his model for the social sciences, supplementing what he learned about scientific method in his philosophy classes with a critical study of the works of Auguste Comte, John Stuart Mill, and Charles Renouvier.

Recently discovered student notes taken in Durkheim's Lycée philosophy classes reveal that from the very beginning of his career he taught that *hypotheses* are absolutely necessary to scientific inquiry. Laws do not leap to the eye, he said. Drawing on examples from the work of scientists such as Galileo, Isaac Newton, and Blaise Pascal, he argued that creative imagination is needed to find explanatory hypotheses. However, there must be some constraints on imaginative hypotheses: They must be simple and precise, explain the known facts, and lead to successful predictions of new facts. He cautioned that scientific laws never lose their hypothetical status no matter how many supporting facts are found, but they can be refuted by a single contradictory phenomenon.

Durkheim's Goals and Methods

Durkheim's philosophy of sociology, which has been called *social realism,* is summed up in his slogan "Treat social facts as things." That is, social facts are real entities with their own underlying causes, not reducible to individual psychological facts. For Durkheim, social reality ultimately consists of collective representations that are shared by members of society, make up the collective or social consciousness, and give rise to social forces and currents. He regarded these social forces as just as real as physical forces.

Social forces cannot be studied through introspection, which reveals only effects and not their causes. One must instead use the experimental method of the natural sciences, which consists of testing hypotheses by comparing their implications with observable facts. In sociology, these included phenomena such as crime statistics, codes of law, religious rites, and various social constraints on ways of thinking and acting. Durkheim's methodology does not quite conform to our contemporary notion of hypothetico-deductivism. He seemed to think that effects could be deduced solely from their causes and that logical relationships were the same as causal relationships. There is no discussion of the distinction between general laws and statements of initial conditions or of the role of auxiliary assumptions in drawing test implications. Nevertheless, he made such assumptions in drawing implications from both his own hypotheses and those against which he was comparing them.

In *The Rules of Sociological Method* (1895), Durkheim championed the use of induction in the social sciences, criticizing economists and others for relying on the abstract, deductive methods of mathematics and philosophy. The method of induction he recommended was not the simple one of enumerating particular observations and then generalizing a conclusion from them. Durkheim was opposed to the mere accumulation of facts, such as one often finds in ethnographies, arguing that science is interested only in what Francis Bacon called "decisive" or "crucial facts," that is, facts that are critical for choosing between competing hypotheses. He thought J. S. Mill's methods of eliminative induction, so called because they are methods for eliminating causal hypotheses that disagree with observations, captured the logic of these crucial experiments.

According to Durkheim, the most useful of Mill's methods for the social sciences is the method of concomitant variation, in which a connection is shown between changes in two different phenomena. For instance, in *The Division of Labor in Society* (1893), Durkheim sought to show that changes in the relative proportions of restitutive and repressive laws vary with the level of complexity of social organization. Although the presence of a concomitance will not alone prove a causal relationship, Durkheim's arguments in *Suicide* (1897) reveal that he thought the *absence* of a concomitance could disprove a causal hypothesis. When there is a concomitance that agrees with our hypothesis, we must still make further comparisons, preferably with other societies, to make sure that it was not just an accident. For example, if we were to find an inverse relationship between the size of families and suicide rates,

we should investigate whether this also holds in countries other than our own. *Suicide* adds two further conditions on the acceptability of a causal hypothesis: The facts must not allow for some alternative explanation, and the hypothesis must not be contradicted by other facts.

Durkheim's Working Method

Durkheim's working method begins with formulating a provisional *definition* of some class of social facts, such as crime, suicide, or religion, in terms of their observable characteristics. This definition is then analyzed into its simplest parts, which often involves tracing what he took to be the historical development of this social phenomenon from its most primitive beginnings. He would then seek the underlying *causes* that would explain these observable characteristics, comparing his own hypothesis with those of others and arguing that his had the greater explanatory scope.

The best example of his use of this method is provided by *The Elementary Forms of Religious Life* (1912). This book begins with a provisional definition of religion, expressed in terms of a distinction between the sacred and the profane, which is maintained by an institution such as a church. It then turns to what Durkheim regarded as the simplest, earliest form of religion: totemism, particularly as it is practiced by the indigenous peoples of Australia. He defended the hypothesis that totemism is the earliest form of religion against the alternatives that either animism or naturism were, arguing that only the totemistic hypothesis could explain the origins of the distinction between the sacred and the profane, as well as the ideas of souls, sacred beings, and religious forces. The totemistic hypothesis explains the sacred character attached to certain totemic objects in terms of the "social forces" that are experienced in periods of collective effervescence that are produced during religious rites. Durkheim then generalized his conclusion to include contemporary religion, maintaining that the feelings of well-being to which it gives rise are but the result of social forces. He held that there is no more reason to think that the reality underlying our religious experiences resembles these experiences than that the reality underlying our experiences of heat, light, and color resembles them in any way.

Durkheim also expanded the concept of a *crucial experiment,* which he also called a "well-made experiment," in *The Elementary Forms*. Anticipating the objection that he had generalized from too few cases, he argued that his study of Australian religions is a well-made experiment. It constituted a crucial test between his totemistic hypothesis, according to which religion rests on social causes, and the animistic and naturistic hypotheses, according to which religion rests on individual, psychological experiences. Although Durkheim generally placed more trust in disconfirmation than in confirmation, here he added that a disconfirming instance must allow for only one possible interpretation and that it cannot be explained without rejecting the hypothesis it seems to disconfirm. He also included among "decisive facts" not only those that contradict a hypothesis but those that a hypothesis fails to explain, either at all or without generating additional problems or making ad hoc assumptions. In short, he included nonrefuting anomalies in the method of crucial experiments.

Durkheim argued that when a law has been established by a well-made experiment, it may be considered a universal law, since he thought that the same effect could not result from different causes. Nevertheless, he maintained that the results achieved by science are always provisional and that the best that science can achieve is probabilities, not certainties.

Explanation

Durkheim sought *causal explanations* of social or collective phenomena, rather than interpretations of the actions of individual social agents. For instance, he was more interested in why suicide rates varied with one's religion or nation than with any individual person's reasons for killing herself or himself. His concept of causal explanation combines the traditional idea of finding the *underlying real essences* behind the appearances with the more modern notion of discovering the laws that govern the phenomena. For Durkheim, the hidden essences of social phenomena both logically entail and causally explain them.

In Durkheim's mind, causes were not simply the antecedents of a general expression in conditional form. He rejected Mill's doctrine of the plurality of causes for the same effect, as well as the plurality of effects for the same cause, at least in the normal state of affairs. Causes for Durkheim are both necessary

and sufficient for their effects, and vice versa. Thus, causes can easily be confused with effects when interpreting some of his claims in his empirical works. Also, since the *functions* of social facts are to be found among their effects, causes can easily be confused with functions, in spite of his warnings in *The Rules* about keeping causal explanations distinct from functional explanations.

For instance, in *Division,* Durkheim presented the rise of cities and towns as both a necessary and a sufficient condition for the increase in social or moral density, a measure of the quantity of interpersonal relationships. Thus, it is unclear whether he meant that an increase in the density of interpersonal relationships is caused by urbanization and thus by population density or whether an increase in social density leads to the creation of towns and villages. Also, it is not clear whether he claimed that specialization is caused by an increase in population density or that specialization is an adaptation to conditions in the social and physical environment. The latter interpretation gains support from Durkheim's making an analogy with Charles Darwin's principle of the divergence of character in *On the Origin of Species* (1859), according to which specialization in nature is not *caused* by the struggle for existence but is rather a solution to it that decreases the competition for resources. In this analogy, the specialization of human labor would be a solution to the Darwinian struggle for existence in a dense population. Durkheim explained that population density alone could simply lead people to emigrate. Specialization allows more people to survive in the same environment, so they do not have to break social bonds that are important to them by emigrating. However, this Darwinian account is a functional, not a causal, explanation of the division of labor. It is not clear whether Durkheim ever did provide a *cause* for the division of labor, leaving himself open to the interpretation that it is caused by physical population density.

Durkheim introduced the distinction between "normal" and "pathological" cases in *Division* in order to deal with anomalies in his theory that labor specialization in turn gives rise to organic solidarity, and thus with exceptions to his one cause/one effect rule. He also made use of this distinction in his attempts to provide a naturalized, sociological basis for ethics, identifying the normal with the average. But, of course, the average is not necessarily morally good, and deviations from the norm in this sense are not necessarily bad.

Durkheim's Antireductionism

According to Durkheim, any time one gives a psychological explanation of a social phenomenon, the explanation is false. Social causes are distinct from psychological causes, and social functions are distinct from individual goals and purposes. He was particularly opposed to intentional explanations, arguing that the fact that some individual or presumably even a group of individuals desired some state of affairs cannot explain what brought it about.

An important goal for Durkheim in writing *Suicide* was to demonstrate that sociology is an autonomous science that does not depend on psychology. According to Durkheim, *societal facts,* such as the varying rates of suicide, dependent on things such as one's country, region, religion, and marital status, cannot be explained simply by summing up individual factors. This is not to deny that there are psychological mechanisms that mediate the effects of society on the individual. For instance, Durkheim maintained that men with large families were more socially integrated and thus less likely to commit suicide than bachelors. Assuming this were true, it would be because people *know* about their familial relations and obligations. But again, explaining the actions of individuals was not Durkheim's concern.

Objections to Durkheim's methodological collectivism (or holism) have often turned on the assumption that it must entail the existence of some sort of entity that exists over and above the individual members of society. Because many of his critics understood his term *collective consciousness* to refer to some such "group mind," he eventually dropped this term and simply referred to a culture's "collective representations" instead. He nevertheless continued to insist that collective representations constituted a realm of real entities distinct from individual mental representations and gave that as the reason why sociological explanations could not be reduced to psychological explanations.

Durkheim did not quite see that the view that sociology does not reduce to psychology no more entails a metaphysical position than any other antireductionist stance in the philosophy of science. For instance, to say that biology does not entirely reduce to chemistry is not to assert the existence of vital spirits. It is simply

to say that there are concepts in biology, such as adaptation, that cannot be defined in chemical terms and generalizations in biology, such as Darwin's theory of evolution through natural selection, that cannot be derived from chemical laws. Similarly, to say that there are concepts, such as social density or social solidarity, that cannot be defined in psychological terms or that cultural preferences for types of suicide or murder weapons cannot be explained psychologically is not to commit oneself to any particular metaphysics.

Lectures on Pragmatism

Durkheim's nephew and collaborator Marcel Mauss regarded Durkheim's series of lectures on pragmatism, which he gave toward the end of his life, as his most important philosophical work. In them, Durkheim defended a realist stance against William James's pragmatist theory of truth. But perhaps his more philosophically interesting argument is his critique of Jamesian pragmatism for being too individualistic. Durkheim argued that truth and reason as well as morality are normative concepts and that one cannot explain normativity from an individualistic perspective, maintaining that it could not be constructed in the life experiences of a single individual. Durkheim's position here calls to mind the Wittgensteinian argument that unless an individual is embedded in society, there is no one to tell her that she is going about things the wrong way, that is, that she is transgressing a norm. However, to say that normativity is a collective or social phenomenon is only to provide a necessary condition for it. Durkheim struggled during his entire career to give a sufficient account of *normativity,* leaving behind the unfinished *La Morale* when he passed away.

Conclusion

Durkheim's works provide an exemplar of the methodological collectivist (holist) position in the philosophy of social science. He argued that *there are social facts that are distinct from individual psychological facts,* such as crime rates, and that require uniquely *sociological* explanations. The social character of normativity comes out most clearly in his critique of the work of William James.

He also maintained some rather sophisticated views concerning comparative theory evaluation, expanding the concept of a crucial experiment to include a role for nonrefuting anomalies as well as for conceptual problems, such as a failure to explain the phenomena without making ad hoc hypotheses.

Warren Schmaus

See also Causation in the Social Sciences; Explanation, Theories of; Explanation Versus Understanding; Holism, in the Social Sciences; Induction and Confirmation; Mill and the Moral Sciences; Realism and Anti-Realism in the Social Sciences; Reductionism in the Social Sciences; Scientific Method; Structural Functionalism, in Social Theory

Further Readings

Durkheim, É. (1951). *Suicide: A study in sociology* (J. A. Spaulding & G. Simpson, Trans.). New York, NY: Free Press. (Original work published 1897)

Durkheim, É. (1978). *Emile Durkheim on institutional analysis* (M. Traugott, Ed. & Trans.). Chicago, IL: University of Chicago Press. (Original works published 1885–1920)

Durkheim, É. (1982). *The rules of sociological method and selected texts on sociology and its method* (W. D. Halls, Trans.). New York, NY: Free Press. (Original work published 1895)

Durkheim, É. (1983). *Pragmatism and sociology* (J. C. Whitehouse, Trans.). Cambridge, England: Cambridge University Press. (Original French publication 1955)

Durkheim, É. (1984). *The division of labor in society* (W. D. Halls, Trans.). New York, NY: Free Press. (Original work published 1893)

Durkheim, É. (1995). *The elementary forms of religious life* (K. Fields, Trans.). New York, NY: Free Press. (Original work published 1912)

Durkheim, É. (2004). *Durkheim's philosophy lectures: Notes from the Lycée de Sens Course, 1883–1884* (N. Gross & R. A. Jones, Trans.). Cambridge, England: Cambridge University Press.

Lukes, S. (1973). *Émile Durkheim: His life and work*. New York, NY: Penguin.

Pickering, W. S. F. (Ed.). (2001). *Émile Durkheim: Critical assessments of leading sociologists* (Third Series). London, England: Routledge.

Schmaus, W. (1994). *Durkheim's philosophy of science and the sociology of knowledge*. Chicago, IL: University of Chicago Press.

Schmaus, W. (2004). *Rethinking Durkheim and his tradition*. New York, NY: Cambridge University Press.

Turner, S. P. (1986). *The search for a methodology of social science: Durkheim, Weber, and the nineteenth-century problem of cause, probability, and action*. Dordrecht, Netherlands: Reidel.

ECONOMETRICS: METHODOLOGICAL ISSUES

This entry reviews the major methodological debates that shaped the field of econometrics by unifying statistics and economics and explains the roles of causality (and probability) and of modeling in econometrics.

In the 1920s, Ragnar Frisch coined the term *econometrics* as the unification of statistics, economic theory, and mathematics to turn economics into a science. Because most current economic theories are expressed in a mathematical language, the constituent mathematics is not considered anymore to be part of modern econometrics. The aim of econometrics is generally conceived as Trygve Haavelmo defined it to be, namely, *a conjunction of economic theory and measurement,* using the theory and technique of *statistical inference* as a bridge pier. But current econometricians emphasize that economies are so highly dimensional, nonstationary, and complicated that economic theory can never precisely specify the underlying process. So a commonly held view today is that econometrics is synonymous with economic statistics.

Methodological Debates

Since its origin in the 1930s, subsequent debates shaped the ideas of what econometrics entails, or is supposed to entail.

The Keynes-Tinbergen Debate

The first two macroeconometric models were constructed by Jan Tinbergen in the late 1930s. The second model was built when Tinbergen was commissioned by the League of Nations to perform statistical tests on business-cycle theories. Tinbergen's new method of econometric testing provoked a great deal of controversy. It was John Maynard Keynes's critique of "Tinbergen's method" that sparked off the debate about the role of econometrics and what it might be able to achieve.

According to Keynes, the technique of multiple correlation analysis that had been adopted by Tinbergen was solely a method for measurement. It contributed nothing in terms of either discovery or criticism. The implication was that if the economic theorist does not provide the modeler with a complete set of causal factors, then the measurement of the other causal factors will be biased. Moreover, Keynes argued that some significant factors in any economy are not capable of measurement or may be interdependent.

The "Measurement Without Theory" Debate

Another early econometric debate in the 1940s started with Tjalling C. Koopmans's book review of *Measuring Business Cycles,* by Arthur F. Burns and Wesley C. Mitchell, published by the National Bureau of Economic Research (NBER). Koopmans's critique was based on Haavelmo's "Probability Approach." In fact, he was defending the Cowles Commission's (CC) structural equation approach

against the NBER's "empiricist position." He accused Burns and Mitchell of trying to measure economic cycles in the absence of any economic theory about the workings of such cycles. According to the CC approach, economic variables are determined by the simultaneous validity of a large number of structural equations describing behavior and technology. Any observed empirical regularity between a number of variables may be the result of the working of several simultaneous structural relations. Because so many empirical relations are valid simultaneously, it may be difficult—or even impossible—to uncover the more fundamental structural relationships. In the absence of experimentation, the identification of these structural relations is possible only if the set of variables involved in each equation, and the manner in which they are combined, is specified by economic theory.

The "Science or Statistical Alchemy" Debate

After 1950, econometrics became a mature field, and the CC approach was the dominant practice. But after two decades (the 1950s and 1960s) of high expectations of econometrics as a producer of reliable predictions and policy advice, in the 1970s, these expectations were increasingly doubted. In the early 1980s, David F. Hendry revisited the Keynes-Tinbergen debate as a backdrop to reiterating the scientific possibilities of econometrics. The ease with which a mechanical application of the econometric method produced spurious correlations suggests alchemy, but according to Hendry, the scientific status of econometrics can be regained by showing that such deceptions are testable. He, therefore, came up with the following simple methodology: "The three golden rules of econometrics are test, test and test."

Edward Leamer's article "Let's Take the Con Out of Econometrics," also written in the early 1980s, is very much about the "myth" of science that empirical research is (randomized controlled) experimentation and scientific inference is objective and free of personal prejudice. The problem of nonexperimental settings (usually the case in economics) compared with experimental settings (common in science) is that the specification uncertainty in many experimental settings may be very small, but this hardly is the case in nonexperimental settings. The often used image that econometrics is like agricultural experimentation (randomized controlled experimentation) is, according to Leamer, grossly misleading.

Econometrics and Causality

Econometrics is a discipline that has paid much attention to how laws, probabilities, and causes fit together. See, for example, the works of T. Haavelmo and Herbert Simon and, more recently, those of Clive Granger and James Heckman. Moreover, it is with econometrics that probabilities entered economics. So econometrics, particularly the CC structural approach, is for some philosophers of science, like Nancy Cartwright and Kevin Hoover, a good starting place for a study of the relationship between causes and regularities.

Econometrics and Empirical Modeling

Tinbergen's econometric models originated a new practice of empirical research in economics, namely, that of *empirical modeling*. Studies of this empirical practice have had an important influence on the development of alternative accounts of models to the more dominant, semantic view. The semantic view is that a model is an interpretation of a theory in which all the axioms of that theory are true. The problem with this view is that econometric models are representations of economic systems of such complexity that these representations cannot be reduced to a small set of axioms. It appeared that models in econometrics function more like instruments, mediating between theories and data in a much more autonomous way.

Marcel Boumans

See also Causation in the Social Sciences; Mathematical Models, Use in the Social Sciences; Models in Social Science; Philosophy of Economics, History of; Probability

Further Readings

Boumans, M. (2005). *How economists model the world into numbers*. London, England: Routledge.

Burns, A. F., & Mitchell, W. C. (1946). *Measuring business cycles*. New York, NY: National Bureau of Economic Research.

Cartwright, N. (1994). *Nature's capacities and their measurement*. Oxford, England: Oxford University Press.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37*(3), 424–438.
Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica, 12*(Suppl.), S1–S115.
Heckman, J. J. (2005). The scientific model of causality. *Sociological Methodology, 35,* 1–97.
Hendry, D. F. (1980). Econometrics: Alchemy or science? *Economica, 47*(188), 387–406.
Hoover, K. D. (2001). *Causality in macroeconomics.* Cambridge, England: Cambridge University Press.
Hoover, K. D. (2006). The methodology of econometrics. In T. C. Mills & K. Patterson (Eds.), *Palgrave handbook of econometrics: Vol. 1. Econometric theory* (pp. 61–87). Basingstoke, England: Palgrave-Macmillan.
Keynes, J. M. (1939). Professor Tinbergen's method. *Economic Journal, 49,* 558–568.
Koopmans, T. C. (1947). Measurement without theory. *Review of Economic Statistics, 29*(3), 161–172.
Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review, 73*(1), 31–43.
Simon, H. A. (1953). Causal ordering and identifiability. In W. C. Hood & J. C. Koopmans (Eds.), *Studies in econometric method* (pp. 49–74). New York, NY: Wiley.
Stigum, B. P. (2003). *Econometrics and the philosophy of economics: Theory–data confrontations in economics.* Princeton, NJ: Princeton University Press.
Tinbergen, J. (1939). *Statistical testing of business-cycle theories* (2 vols.). Geneva, Switzerland: League of Nations.

Economic Anthropology

Economic anthropology arose where the disciplines of anthropology and economics meet. The Greek *oikonomia* meant household management with the aim of self-sufficiency through thrift, careful budgeting, and the avoidance of trade. This ideal persisted in Europe up to around 1800. Adam Smith's revolution focused on the division of labor and the functioning of markets in the wider society. *Economy* has since come to refer primarily to market exchange, although the old meaning of making ends meet persists. Economy could still mean "putting one's house in order," both practically and figuratively. Everyone should feel "at home" in a world made by markets, but we cannot survive on the basis of market economy alone.

Two Pioneers

The sociologist Marcel Mauss asked, "What rule compels the gift to be reciprocated?" People find the personal character of the gift compelling, since it evokes diffuse social and spiritual ties. Human institutions everywhere are founded on the unity of the individual and society, freedom and obligation, and self-interest and concern for others. Modern capitalism and economics rest on an unsustainable attachment to one extreme. Mauss held that markets and money are universal, though not in their current impersonal form, while advocating a pragmatic approach to the human economy of relevance to people's daily lives.

The political economist and philosopher Karl Polanyi's *The Great Transformation* showed how a free-wage labor market in Victorian England led to the economy becoming "disembedded." The 20th century's crises and World Wars were the result. He later argued that the "formal" and "substantive" meanings of the word *economy* had been conflated. The first refers to economizing, a means–end relationship, whereas the second is concerned with the provisioning of material wants. The economic institutions of preindustrial societies guarantee social survival, whereas the abstract market principles that drive the economy of industrial societies do not. This led him to abandon the study of modern economies to the economists.

For both these thinkers, society is founded on a limited number of economic principles that are distributed widely in history but combined variably. Like Karl Marx, they rejected the reduction of society to capitalist markets. They saw the economy as being pulled in two directions at once: inward to secure local guarantees of a community's rights and interests and outward to make good deficiencies of local supply by engaging with foreigners through money and markets. An economic anthropology for the 21st century would perhaps do well to build on their example.

The History of Economic Anthropology

The field's origins lie in the democratic revolutions of the 18th century. Philosophers like Jean-Jacques Rousseau offered a revolutionary critique of the

premise of inequality and constructive proposals for a more equal future. Karl Marx and Friedrich Engels subsequently made fertile use of this precedent in their own critique of the state and capitalism. Few writers have come close to matching Marx's vision of economic history as a whole.

Modern anthropologists asked whether the economic behavior of "savages" was as efficient and rational as the Western equivalent. Ethnographers sought to engage the general propositions of "neoclassical" economics with their particular findings about "primitive societies." They failed mainly because they misunderstood the economists' premises.

In midcentury, anthropologists argued about the theories and methods needed to study tribesmen and peasants. "Formalists," who saw an abstract individualism everywhere, held that the tools of mainstream economics were adequate to the task, while "substantivists" claimed that institutional approaches were more appropriate. This "formalist versus substantivist" debate ended in a stalemate, opening the way for Marxists and feminists to exercise a brief dominance, but they too at first drew mainly on exotic subject matter.

Developments From the 1980s

Anthropologists now address the full range of economic organization, but so far, they have preferred to stick with ethnographic observation. Formalism lives on as the "new institutional economics." This consists in extending market models and rational choice approaches into new areas, while relying on the concept of *transaction costs*. Many economic anthropologists have tried to open up the black box of "culture." The best combine an ethnographic sensitivity to the "social life of things" with the recognition that cultural valuation is itself shaped by inequalities of wealth, power, and status. This "cultural turn" has extended the substantivists' critique of the bourgeois economic categories by showing that they offer just another local model and an unattractive one at that.

We all seem to be living in a world unified by capitalism, so economic anthropologists now investigate that. A significant strand studies the material culture of consumption at home. There has been some excellent ethnography of industrial work, and anthropologists have engaged critically with the great questions of modern history: unequal development, socialism, and global capitalism. The aim of humanizing the anonymous forces governing our lives has produced a fashion for the anthropology of money and financial institutions. This field has exploded since 2008. It is hard to argue now that economies prosper only if markets are freed from political bondage. So there may be more scope for alternative approaches to the economy.

The basic issue remains whether the forms of market economy that have dominated for two centuries rest on principles of universal human validity. A new synthesis of anthropology, world history, and economics might provide some answers.

Keith Hart

See also Economic Sociology; Ethnography, Philosophical Aspects of; Homo Economicus; Individualism, Methodological; Markets and Economic Theory; Money; Social Anthropology

Further Readings

Carrier, J. (Ed.). (2012). *A handbook of economic anthropology* (2nd ed.). Cheltenham, England: Edward Elgar.

Gudeman, S. (2008). *Economy's tension: The dialectics of community and market.* Oxford, England: Berghahn Books.

Hann, C., & Hart, K. (Eds.). (2009). *Market and society: The great transformation today.* Cambridge, England: Cambridge University Press.

Hann, C., & Hart, K. (2011). *Economic anthropology: History, ethnography, critique.* Cambridge, England: Polity.

Hart, K. (2000). *The memory bank: Money in an unequal world.* London, England: Profile. (Republished as *Money in an unequal world,* 2001, New York, NY: Texere)

Hart, K., Laville, J., & Cattani, A. D. (Eds.). (2010). *The human economy: A citizen's guide.* Cambridge, England: Polity.

Mauss, M. (1990). *The gift: The form and reason for exchange in archaic societies.* London, England: Routledge. (Original work published 1925)

Polanyi, K. (2001). *The great transformation: The political and economic origins of our times.* Boston, MA: Beacon Press. (Original work published 1944)

Polanyi, K., Arensberg, C., & Pearson, H. (Eds.). (1957). *Trade and market in the early empires: Economies in history and theory.* Glencoe, IL: Free Press.

Economic Sociology

Economic sociology can be defined as the sociological analysis of economic phenomena. It was created during the 19th century and is currently one of the most popular subfields of sociology in the United States and elsewhere. Its two most important founders are Karl Marx and Max Weber. The field revived in the 1980s and operates today with concepts such as *embeddedness, social capital,* and *field*.

The Classics and Onward

Karl Marx created his own kind of analysis that does not fit very well into any modern social science discipline. Max Weber, in contrast, helped make sociology accepted as a university subject and was extremely interested in *Wirtschaftssoziologie*. Like Marx, Weber was very well versed in economic theory as well as economic history, and both of these disciplines helped him formulate a powerful program in economic sociology. This program can be found in Chapter 3 of his *Economy and Society,* titled "Sociological Categories of Economic Action."

Like all of Weber's sociology, his economic sociology was part of what he called "interpretive sociology." In this type of sociology, it is imperative that the actor's view of things is part of the analysis. This perspective also informs Weber's earlier works, written before he decided to call himself a sociologist. This is, for example, true for the classic *The Protestant Ethic and the Spirit of Capitalism.*

Weber died in 1920, and during the period from 1920 to the 1980s, economic sociology more or less disappeared from the universities. Very few courses were taught in this field, and very few studies were carried out. Despite this fact, three major economic sociologists were active during this period. They are Joseph Schumpeter, Karl Polanyi, and Talcott Parsons.

Schumpeter wrote a series of essays in economic sociology but was mainly seen as an economist. His most important essay in economic sociology is on the role of taxation in society and can be described as an early and vigorous argument for a fiscal sociology. Many people also think that Schumpeter's economic analysis was quite sociological in nature. This is especially true of *Capitalism, Socialism and Democracy.*

Karl Polanyi can be described as an interdisciplinary and normatively engaged scholar. His work has had a great impact on economic sociology, not the least on people critical of the existing economic order. According to Polanyi, the economy had originally been subordinate to religion and politics but was set free from all restraints in 19th-century England. This resulted in huge and destructive changes in human society, analyzed and recorded in Polanyi's masterpiece *The Great Transformation*. Polanyi was also a skillful theoretician, and his notion that all economic acts can be characterized as a form of reciprocity, redistribution, or exchange is generally accepted in modern economic sociology.

Talcott Parsons had (like Weber) been educated as an economist, and he later put this knowledge into use when he became a sociologist. Together with his student Neil Smelser, Parsons published a famous study centered on the idea that the economy can be seen as a social system. Just like any social system, Parsons and Smelser argued, an economy has to deal with issues such as adaptation, setting goals, and so on.

Modern Economic Sociology

The birth of modern economic sociology is usually associated with the publication in 1985 of a famous article by Mark Granovetter on *embeddedness*. In contrast to Polanyi, who used this term in a vague sense, Granovetter specified that embeddedness means that economic action is always embedded in ongoing *networks* of social relations. The idea of "homo economicus" is not so much wrong from this perspective because economic man is seen as exclusively driven by profit or wealth but because it ignores the role of social structure.

In the two to three decades after Granovetter's article, "new economic sociology" (as it is sometimes called) has advanced very quickly. Today, the subject is totally institutionalized in the sense that most major sociology departments routinely teach courses in economic sociology, there are several textbooks and handbooks on economic sociology, and most national sociological associations have sections on economic sociology.

Economic sociologists have made important contributions to a number of areas. They have, for example, closely followed the transition from the traditional firm to the so-called shareholding view of the firm. The role that networks play in people getting

jobs and in firms collaborating with one another has similarly been carefully studied. Another important area has to do with the role of gender differences in the economy. More generally, there exist today studies on pretty much everything, from how to buy a house in France to how the global financial services firm Lehman Brothers collapsed in the fall of 2008.

In carrying out their studies, economic sociologists draw on a wide variety of methods. Some economic sociologists conduct surveys or work with huge data sets. Others conduct participant observation studies, apply a cultural approach to what they study, or use a historical and comparative perspective. Like modern sociology, modern economic sociology is pluralistic. There exist economic sociologists who are very close to economists and also those who refuse to accept anything that economists do.

When economic sociology was revived in the mid-1980s, it was hoped that economic sociology would one day be strong enough to take on economists. Alternatively, it was hoped that economists would become interested in the sociological analysis of the economy. To date, none of these things have happened. Nonetheless, today's economic sociology does constitute a lively and interesting area of study.

Richard Swedberg

See also Homo Economicus; Markets and Economic Theory; Philosophy of Sociology, History of; Social Capital; Social Networks; Weber and Social Science: Methodological Precepts; Weber's *Verstehende* Approach

Further Readings

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology, 91,* 481–510.

Parsons, T., & Smelser, N. (1956). *Economy and society: A study in the integration of economic and social theory.* New York, NY: Free Press.

Polanyi, K. (1957). *The great transformation.* Boston, MA: Beacon. (Original work published 1944)

Polanyi, K., Arensberg, C., & Pearson, H. (Eds.). (1971). *Trade and market in the early empires.* Chicago, IL: Regnery. (Original work published 1957)

Schumpeter, J. A. (1994). *Capitalism, socialism and democracy.* London, England: Routledge. (Original work published 1942)

Smelser, N., & Swedberg, R. (Eds.). (2003). *The handbook of economic sociology* (2nd rev. ed.). New York, NY: Russell Sage Foundation.

Swedberg, R. (2003). *Principles of economic sociology.* Princeton, NJ: Princeton University Press.

Weber, M. (1958). *The Protestant ethic and the spirit of capitalism.* New York, NY: Charles Scribner. (Original work published 1904–1905)

Weber, M. (1978). *Economy and society: An outline of interpretive sociology* (2 vols.). Berkeley: University of California Press. (Original work published 1922 in German as *Wirtschaft und Gesellschaft: Eine Übersicht der verstehenden Soziologie,* Tubingen, Germany: Mohr)

Economics of Scientific Knowledge

This entry introduces a blossoming research field in the philosophy of science in which certain tools used by economists are applied to the analysis of the production and evaluation of scientific knowledge. On this view, scientists are seen as economic agents. The entry reviews the different approaches this field has taken. The special interest of this field lies in the way philosophy makes use of the tools and models of a particular social science, economics, in analyzing scientific knowledge and the way scientists act in evaluating their theory production.

Background

The term *economics of scientific knowledge* (hereafter ESK) was coined in the 1990s as a reaction to the field known from the 1970s as the sociology of scientific knowledge. The latter had been defined by the members of the so-called Strong Program in contraposition to the classical notion of a sociology of science, referring to the distinction between the sociological explanation of the institutional, political, and cultural aspects of science, on the one hand, and the sociological explanation of the *cognitive* aspects of science, on the other. The sociology of science would be devoted to the external (nonepistemic) aspects of science, whereas the sociology of scientific knowledge would study the *internal* content of science, that is, why certain theories, facts, or paradigms are accepted or rejected. Sociologists in the Strong Program derived some radically relativist conclusions from this starting point, in opposition to most traditional views about scientific knowledge.

As a result of this, an open question remained, namely, whether the application of analytical instruments drawn from the economist's toolkit to the understanding of the process of knowledge generation —that is, the view of scientists as agents within an economic model—would support the relativist claims of radical sociologists. Or, on the contrary, it would serve to "save" the intuitive character of scientific knowledge as a paradigm of "objectivity." As this entry will show, most contributions to ESK fall under the second of these options.

In retrospect, we can realize that many prior works can be identified under the ESK label, though the main production of papers and books on the topic has occurred in the past two decades. One possible way of classifying all these works is according to the type of economic models or metaphors they attempt to apply to the study of the creation of scientific knowledge. From this point of view, we can distinguish *formal* (or "mathematical") from *nonformal* (or "institutional") approaches.

Formal Approaches

One of the most important contributions in the first group is Philip Kitcher's 1990 article titled "The Division of Cognitive Labor," later reprinted in an extended form as the last chapter of his 1993 book *The Advancement of Science,* in which he develops a set of models based on the assumption of interacting, rational, self-interested scientists. According to Kitcher, the aim of his models is to "identify the properties of epistemically well-designed social systems," that is, to study how a group of individuals, by interacting with each other, manage to reach a consensus gradually, working under certain rules. Other interesting mathematical models of scientific activity that have been developed during the past decades refer to a tit-for-tat "game" between researchers and journal editors, to the way in which researchers try to change the subjective probabilities of their colleagues, or to the decision whether to replicate another researcher's experiments, or the decision of accepting a more "popular" theory or defending a more heterodox one, on the basis of the different information about both theories each individual scientist has. The last two cases show the possible existence of more than one equilibrium in the "cognitive state" of the scientific community—which can lead to phenomena of path dependence, inefficiency, and sudden "revolutions." Some more recent contributions have analyzed the properties of the priority rule and the choice of methodological rules, while Jesús Zamora-Bonilla has studied the negotiation taking place with regard to the interpretation of empirical findings. More recently, due to the availability of more powerful software, the use of simulation models to study Kitcherian "division of epistemic labor" problems has become relatively common, especially in cases where complexity is relevant. Many of these articles are grounded on the simulation models of Reiner Hegselmann and Urich Krause, which in turn are inspired by the work of Keith Lehrer and Carl Wagner on "rational" belief aggregation.

A brief review of the formal approach to ESK would be incomplete without mention of Samir Okasha's 2011 article, "Theory Choice and Social Choice." Here, an analogy is suggested between the aggregation of individual preference functions, to which the famous Arrow impossibility theorem was originally applied (showing that there is no way of constructing a "social" preference function that respects certain minimal and reasonable requirements), and the combination of different "scientific values" that would result in something like an objective "epistemic preference function." Okasha argues that this analogy justifies in a way Thomas Kuhn's thesis that there is no algorithm allowing us to determine in an objective sense when one theory is epistemically better than another; the difference would be that whereas Kuhn's original intuition, and almost all of the subsequent interpretations, has held that the problem is that there are an indefinite number of possible ways of aggregating the different scientific values, none of them being justifiably better than the others, Arrow's theorem would imply that the problem is that there is simply no "rational' way of performing such an aggregation. It can be argued, however, that even if it is true that several scientific values cannot be algorithmically combined, scientists might agree in employing a common epistemic scale that is nonoptimal for each one but that is an equilibrium in a negotiation (or "social contract") about how to evaluate research.

We must not forget what is very likely the first contribution to ESK. The 19th-century American pragmatist philosopher Charles S. Peirce made an amazing application of the then just-invented marginal analysis of economic functions to the process of deciding how much effort to expend in alternative

research projects. This "cost–benefit" approach has been further explored by other authors, such as Nicholas Rescher and Gerhard Radnitzky.

Most of the works referred to in the previous paragraphs support the conclusion that, in spite of being motivated by "personal" goals like recognition, power, or "credit" (to use the concept coined by Pierre Bourdieu, whose sociological insights have been decisively influential and inspiring in the development of ESK), we can reasonably expect that, under certain circumstances and mechanisms of interaction, a group of scientists could attain results that will score high on an epistemic scale.

Nonformal Approaches

With respect to the nonformal approaches to ESK, this entry identifies them as "institutional," in line with the tradition in institutional economics, which is basically qualitative in the sense of grounding its conclusions not on logico-mathematical theorems or models but on mere commonsense reflection on our knowledge of the situations and of the available data. In this area, the basic motivation has been to discuss whether science can be legitimately understood as a "marketplace of ideas," very likely having in mind the conclusion that if science is a kind of free market, then it can be led, "as if by an invisible hand," toward the truth, in spite of the base motivation of single scientists. A pioneering work on this idea was Michael Polanyi's article "The Republic of Science." Some more recent works that share in some way the "Austrian economics flavor" of Polanyi's are that of the philosopher of biology David Hull (though this is hardly classifiable as belonging to any "economic" approach, being more biologically inspired) and those of Yanfei Shi and Alan Walstad. Christoph Lütge offers a similar view but one that is based on James Buchanan's constitutional political economy (or "public choice" theory), which has also influenced the work of Shi.

There are also authors who have employed the market metaphor but in a critical sense, either trying to illuminate in what respects science is not a market or trying to show that science's being something like a market makes it inefficient in the pursuit of truth. In the second group, we can include the work of some sociologists of science, such as Bruno Latour and Steve Woolgar, who employ a Marxist notion of "market" and also some works in which it is argued that economic science, in particular, does not function as if guided by an invisible hand toward the truth but is governed by other forces and social mechanisms that do not warrant that epistemically efficient theories are unanimously accepted.

Jesús P. Zamora-Bonilla

See also Actor-Network Theory; Context of Discovery Versus Context of Justification; Institutional Economics; Invisible Hand Explanations; Kuhn on Scientific Revolutions and Incommensurability; Social Epistemology; Social Studies of Science and Technology; Sociology of Knowledge and Science; Strong Program in the Sociology of Scientific Knowledge

Further Readings

Albert, M. (2006). *Product quality in scientific competition* (Papers on Strategic Interaction 2006–6). Jena, Germany: Max Planck Institute of Economics.

Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information, 14,* 19–47.

Douven, I. (Ed.). (2009). Computer simulations in social epistemology. *Episteme, 6*(2), 107–109.

Hands, D. W. (1994). The sociology of scientific knowledge and economics: Some thoughts on the possibilities. In R. Backhouse (Ed.), *New perspectives in economic methodology* (pp. 75–106). London, England: Routledge.

Kitcher, P. (1990). The division of cognitive labor. *Journal of Philosophy, 87,* 5–22.

Kitcher, P. (1993). *The advancement of science: Science without legend, objectivity without illusions.* Oxford, England: Oxford University Press.

Latour, B., & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts.* Beverly Hills, CA: Sage.

Lehrer, K., & Wagner, C. G. (1981). *Rational consensus in science and society.* Dordrecht, Netherlands: D. Reidel.

Luetge, C. (2004). Economics in philosophy of science: A dismal contribution? *Synthese, 140,* 279–305.

Mäki, U. (2004). Economic epistemology: Hopes and horrors. *Episteme, 1,* 211–222.

Mirowski, P., & Sklivas, S. (1991). Why econometricians don't replicate (although they do reproduce). *Review of Political Economy, 31,* 146–163.

Muldoon, R., & Weisberg, M. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of Science, 76,* 225–252.

Okasha, S. (2011). Theory choice and social choice: Kuhn versus Arrow. *Mind, 120*(477), 83–115.
Peirce, C. S. (1879). *A note on the theory of the economy of research* (United States Coast Survey for the Fiscal Year Ending June 1876). Washington, DC: Superintendent of the United States Coast Survey.
Polanyi, M. (1962). The republic of science: Its political and economic theory. *Minerva, 1,* 54–73.
Radnitzky, G. (1986). Towards an "economic" theory of methodology. *Methodology and Science, 19,* 124–147.
Rescher, N. (1976). Peirce and the economy of research. *Philosophy of Science, 43,* 71–98.
Shi, Y. (2001). *The economics of scientific knowledge: A rational-choice neoinstitutionalist theory of science.* Cheltenham, England: Edward Elgar.
Strevens, M. (2003). The role of the priority rule in science. *Journal of Philosophy, 100,* 55–79.
Walstad. A. (2002). Science as a market process. *Independent Review, 7,* 5–45.
Wible, J. R. (1998). *The economics of science: Methodology and epistemology as if economics really mattered.* London, England: Routledge.
Zamora-Bonilla, J. P. (1999). The elementary economics of scientific consensus. *Theoria, 14,* 461–488.
Zamora-Bonilla, J. P. (2002). Scientific inference and the pursuit of fame: A contractarian approach. *Philosophy of Science, 69,* 300–323.
Zamora-Bonilla, J. P. (2006). Rhetoric, induction, and the free speech dilemma. *Philosophy of Science, 73,* 175–193.

Econophysics

Econophysics (a term coined in 1995 in a conference on statistical physics held at Kolkata, India) views the problems in economics in a physical way and attempts to solve them employing the techniques of physics. Indeed, viewed more generally, econophysics holds that the dynamical aspects of societies, and markets in particular, are purely physical in origin and nature, in contrast to the received view of economics as a typical social science.

Methods of obtaining knowledge toward the truth can be either deductive or inductive. Mathematics is a standard example of deductive knowledge (though not all of it can be deduced from axiomatic logic), not requiring laboratory or natural observation to validate its truths. In contrast, all the natural sciences are essentially inductive in origin, based on observations. The tools of mathematics and logic are only employed in these sciences subsequently in order to find and establish relationships among the observations. These quantifiable links between often seemingly unrelated observations help identify the basic truths of nature. Physicists, chemists, biologists, economists, or sociologists all tend to do the same. In all its various manifestations, inanimate, biological, or sociological, reality perhaps employs the same elegant truth code expressed differently in its various parts. Scientists having different perspectives in mind perceive them differently. The basic truth established, therefore, in one branch of natural science (say physics) should not be seen as invalid in another (say chemistry or biology), and the same should hold true for economics if viewed as another branch of natural science.

Social sciences like economics can be empirical or inductive, like physics or the other natural sciences, where deductions or logical derivations, correlating various empirical observations, play the role indicated above. But that need not be mistaken to be an indication for an essentially deductive science. Just like biophysics or biochemistry (each borrowing established knowledge, i.e., truths from the parent sciences of physics or chemistry, to develop biology), econophysics can help develop economics by utilizing the knowledge borrowed from physics, the most developed branch of natural science. If physics can offer such a helping hand, the indispensability of econophysics can be acknowledged.

Econophysicists' observations that the fluctuation in the size of a company decreases with the size itself, following a power-law, and that the network of companies in the market shrinks in a crisis period have given major insights into market dynamics. On another front, the problem of repetitive games of collective choices within a community of agents, each trying to belong to the minority group (the payoff being greater for those choosing the less crowded option), has been successfully modeled following the dynamical multi-attractor learning models in statistical physics of frustrated many-body systems. Observations regarding the precise nature of the fluctuations (over averages) of stock prices have been extensively analyzed recently and have provided the initial motivation for the formulations of minority game models: the El Farol Bar problem or its generalization, called the Kolkata Paise Restaurant problem. In both cases, the aforementioned techniques provided quite successful solutions.

Among the major research themes in econophysics today, one is to characterize the "natural" forms of economic inequality in society. With the identification of money or wealth exchange dynamics between any two traders as the two-body scattering or collision processes in a gas, considerable progress has recently been made. In particular, the entropy-maximizing stochastic kinetic exchange dynamics, studied extensively in the context of statistical mechanics of gases for more than 100 years, are seen to be equivalent to the (microeconomic) utility maximization problem of the agents in the market (subject to budget constraints). Utilizing this equivalence and the well-established physical models of kinetics in gases, considerable progress has been made. It shows how the socialist norm of zero-dispersion (absolute equality) distribution gets quickly destabilized to the entropy-maximizing, Gibbs-like distribution of money or wealth in such exchange markets. Incorporation of the saving propensity of the traders immediately gives the nonmonotonic, Gamma-like distribution for the bulk of society, with the Pareto power-law decay distribution of the number of superrich in the market with their money holdings or wealth. These studies indicate that inevitable economic inequalities follow, modeled on the analogy with the natural laws of entropy maximization in such dynamical processes. They also indicate ways to control such dispersions in the distribution.

However, the social dynamics of a large collection of individuals with (possibly finite) intelligence placed in a market cannot be trivially identical to the many-body dynamics of inane material particles in a gas or to similar simplistic physical systems. Hence, the approach has attracted criticism and the despair of some about the ultimate future of econophysics. This is not quite unexpected. However, proper improvisations can easily overcome many such problems. Also, current limitations need not pose any fundamental problem to econophysicists: Even intelligent learning models developed in physics or computer science, though immature, are already in use in medical applications and in designing new generations of computers, among other things. Applications of such learning models in multi-agent repetitive games are already leading to intriguing success. Prospects therefore look quite healthy. Though rather young as a discipline (formally started only in 1995) and not quite developed or established yet, econophysics has already infused economic science with a unique natural-scientific spirit and guiding philosophy.

Bikas K. Chakrabarti

See also Biology and the Social Sciences; Complex Networks Theory and Social Phenomena; Mathematical Models, Use in the Social Sciences; Models in Social Science; Multi-Agent Modeling; Reduction and the Unity of Science; Reductionism in the Social Sciences

Further Readings

Chakrabarti, B. K., Chakraborti, A., Chakravarty, S. R., & Chatterjee, A. (in press). *Econophysics of income and wealth distributions*. Cambridge, England: Cambridge University Press.

Mantegna, R. N., & Stanley, H. E. (2007). *Introduction to econophysics: Correlations and complexity in finance*. Cambridge, England: Cambridge University Press.

Sinha, S., Chatterjee, A., Chakraborti, A., & Chakrabarti, B. K. (2010). *Econophysics: An introduction*. Berlin, Germany: Wiley-VCH.

Ego

The notion of ego is central to psychoanalysis and has been a significant concept in psychology too, more generally. It has been a significant notion originating from social science, as opposed to that of the "self" in its predominantly philosophical use. This entry starts by presenting the varying conceptions of the ego put forward by Sigmund Freud, goes on to explain its functions, considers different post-Freudian developments (some anticipating "functionalism" in recent philosophy of mind), and ends by contrasting philosophy with psychoanalysis in this regard. The latter has important implications for the analysis of agency.

Definitions

Some use the term *ego* to refer to the self as a whole; others use it to designate a part of the mind. Freud uses it both ways.

In some of his writings, particularly on narcissism, Freud identifies the ego with the self. He makes this equation explicit in the fourth paragraph of *Civilization and Its Discontents,* where he says that

normally there is nothing of which we are more certain than the feeling of our own self, of our own ego.

In other works, Freud thinks of the ego as just part of the mind. In *The Ego and the Id,* he says that in each individual there is a coherent organization of mental processes, which he calls "the ego." Psychoanalysts sometimes call this concept the "agency concept" because Freud describes this mental organization as the mental agency that supervises its own constituent processes. He also attributes to it special attributes, especially consciousness, and special functions.

In his early writings, Freud theorized that the ego was entirely conscious, but he indicated a change of mind in 1922, when he read a paper at the Seventh International Psycho-Analytical Congress. In the abstract of his paper, Freud notes that previously he had regarded the repressed as coinciding with the unconscious, whereas the ego was identified with the preconscious and conscious, but, later, two facts mandated a change in view. The two facts are (1) resistance proceeding from the ego during analysis and (2) an unconscious sense of guilt. Because of these two facts, Freud argues, the correct view is that the ego is not only partly conscious or preconscious but is also partly unconscious.

Functions

As to the functions of the ego, Freud held that one important function is to deal with the perception of dangers emanating from both the external and the internal world. The internal threats come partly from the id, the totally unconscious part of the human mind that is present at birth, out of which the ego develops and which seeks only the satisfaction of instinctual needs.

The id threatens the ego in two ways, according to Freud. First, an excessive instinctual demand can threaten the dynamic relationship between the id and the ego and change the ego back into being a portion of the id. Second, experience may have taught the ego that satisfaction of some instinctual demand that is not in itself intolerable might nevertheless involve dangers in the external world, so that an instinctual demand of that kind is perceived to be dangerous. The ego, then, is fighting on two fronts: It has to defend its existence against an external world threatening its annihilation, and it must cope with excessive instinctual demands from the id.

There is a third front as well. Prior to age 5, the ego mediates between the id and the external world, guided at all times by the injunctions of a modified pleasure principle. At the end of the fifth year, Freudian theory holds, an important change takes place. A new psychic agency develops out of the ego as people in the external world, mainly the parents, are abandoned as objects and the new agency carries on the functions of the abandoned objects. This agency, which Freud calls the "superego," observes the ego, gives it orders, and judges it, just like the parents whose place it has taken. The superego at times also threatens the ego with punishment, just as the parents have, but, Freud notes, it often displays a severity for which no model has been provided by the real parents. In addition, it calls the ego to account not only for its deeds but also for its thoughts and unexecuted intentions.

When the ego is threatened by the world, the id, or the superego, often reacts by using defense mechanisms, the most important of which is repression. Freud described the essence of repression as simply turning something away and keeping it at a distance from consciousness. The concept of repression plays a key role in Freud's theory of dreams, his theory of neuroses, as well as in his theory of treatment. He called the theory of repression the cornerstone of psychoanalysis.

Other ego defense mechanisms include sublimation, denial, reaction formation, projection, and regression. They operate at an unconscious level, leaving the subject unaware of the true purpose of the defensive behavior.

Another important function of the ego is dream censorship. In one of his best-known works, *The Interpretation of Dreams,* Freud describes how the ego is weakened when it goes to sleep and reacts to the id's forcing upon it infantile repressed wishes. The ego reacts by censoring the unconscious material emanating from the id, allowing the material to enter the preconscious only in a disguised form. If, upon waking, a person remembers a dream, then it is always the manifest content of the dream that is remembered; it is the product of the ego's distortion of the latent and unconscious dream thoughts and images. Some features of the manifest dream, the dream as remembered by the dreamer, may be due not to distortion of infantile repressed wishes but to urges from waking life, which Freud calls the day residue.

Post-Freudian Psychoanalytic Developments

Freud's ambiguous use of the term *ego* is reflected in successor psychoanalytic theories. One of these is the theory of ego psychology, developed by Heinz Hartmann, David Rappaport, Ernst Kris, Merton Gill, and others. The ego psychologists generally follow the view of *The Ego and the Id,* characterizing the ego as an agent having specific functions, including those described by Freud.

In contrast, those in the object relations tradition, developed by Melanie Klein, R. D. Fairbairn, and D. W. Winnicott, follow Freud in his writings on narcissism, identifying the ego with the self and often using the two concepts interchangeably.

One leading object relations theorist, R. D. Fairbairn, speaks of the splitting of the ego, whereas others in the same paradigm talk about the splitting of the self. How can an ego or a self split? The answer lies in how Fairbairn and others use the idea of the internalization of objects.

The term *object* in object relation theories is often, but not consistently, used to refer to persons in an infant's environment, such as the mother or father, or to parts of persons, such as the mother's breast, but when object relation theorists turn to internalized objects, they refer not to external things but to something mental, such as an experience or representation. At some point, the infant "internalizes" the objects in its environment, or more precisely, it internalizes or makes part of itself an *experience* or *representation* of the mother or father, or some other external object. The internalized object, on Fairbairn's account, is divided into a good and a bad object, with the ego repressing the latter. Because part of the ego is thought to remain attached to the repressed object and part of the ego is not so attached, Fairbairn speaks of a splitting of the ego. Others appear to be expressing the same thought when they speak of a splitting of the self.

In still another variant of psychoanalysis, self-psychology, the psychology associated with Heinz Kohut and Otto Kernberg, the ego is generally identified with the entire self, as it is in object relations theory.

Postulation of an ego can give rise to certain mysteries depending on what is postulated. As noted earlier, in *The Ego and the Id,* Freud conceives the ego to be only a part of the mind, yet he also speaks of it as a mental agency that supervises its own processes, discharges excitations into the external world, goes to sleep at night, and censors dreams. To philosophers and psychologists who think of an agent as a person, it is puzzling how a part of a person can also be an agent, something that decides, acts, and goes to sleep.

Another mystery concerns the nature of the ego. Some ego psychologists claim to sidestep questions about its essence by identifying it with its functions, as Freud himself did. The ego, they say, is just that part of the mind that censors dreams and deals with perceived dangers coming from the external world, the id, and the superego by using repression and other defense mechanisms. Here, Freud and the ego psychologists anticipate a widely held view in the philosophy of mind, known as *functionalism.* Mental states, such as beliefs and desires, are said by functionalists in philosophy to be kinds of internal states that stand in certain functional relationships with the agent's behavior and other mental states.

Philosophy Versus Psychoanalysis

There is this difference between the philosophical and the psychoanalytic view. Philosophers characterize functional states in causal terms. A belief is said to be that sort of mental state that causes certain types of behavior and other mental states. Psychoanalysts, however, are referring not to causes but to what the ego does, such as censoring dream content and triggering defense mechanisms, although what it does has certain significant effects.

Psychoanalysts who identify the ego with the self might appear to spare themselves the problems of dealing with the agency concept. After all, as Freud pointed out, there is nothing of which we are more certain than the feeling of our own self. Of course, if Freud is right about one of his central contentions, that there exists in each of us a deep unconscious, there is much about our own self that we know nothing, unless we undergo a successful psychoanalysis; for without psychoanalytic treatment, we have no access to that part of the self that is unconscious.

There is also a philosophical problem about the self, and hence about the ego if we identify the two. As David Hume noted, whenever we introspect, we come across a particular mental state, such as a feeling or thought or sensation, but never the self. How,

then, do we know it even exists in ourselves, let alone in others? Ego psychologists and object relations theorists have various answers to this query, but whether or not their answers are satisfactory, the problem of proving the existence of an unobservable ego arises generally, whether the ego is conceived of as the entire self or as an agent that is just a part of the mind. The problems diminish for nonpsychoanalytic theories when the self is said to be not an entity within a person but the entire person or when it is identified with a collection of mental states and processes.

Edward Erwin

See also Agency; Mind–Body Relation; Psychoanalysis, Philosophical Issues in; Schizophrenia: Psychoanalytic, Phenomenological, and Contemporary Philosophical Approaches; Self-Knowledge

Further Readings

Eagle, M. (1984). *Recent developments in psychoanalysis: A critical evaluation.* New York, NY: McGraw-Hill.

Eagle, M. (2011). *From classical to contemporary psychoanalysis: A critique and integration.* New York, NY: Routledge.

Erwin, E. (1997). The self in psychotherapy. In *Philosophy and psychotherapy* (pp. 35–59). London, England: Sage.

Fairbairn, W. D. (1952). *Psychoanalytic studies of the personality.* London, England: Tavistock/Routledge & Kegan Paul.

Freud, S. (1900). *The interpretation of dreams* (Standard ed., Vols. 4 and 5, pp. 1–627). London, England: Hogarth.

Freud, S. (1914). *On narcissism: An introduction* (Standard ed., Vol. 14, pp. 73–102). London, England: Hogarth.

Freud, S. (1923). *The ego and the id* (Standard ed., Vol. 19, pp. 12–66). London, England: Hogarth.

Freud, S. (1930). *Civilization and its discontents* (Standard ed., Vol. 21). London, England: Hogarth.

Freud, S. (1949). *An outline of psycho-analysis.* New York, NY: W. W. Norton.

Hartmann, H. (1950). Comments on the psychoanalytic theory of the ego. *Psychoanalytic Study of the Child, 5,* 74–96.

Meyer, J., & Bauer, B. (2002). Ego psychology. In E. Erwin (Ed.), *The Freud encyclopedia: Theory, therapy, and culture* (pp. 169–171). New York, NY: Routledge.

Wolf, E. (2002). Self psychology. In E. Erwin (Ed.), *The Freud encyclopedia: Theory, therapy, and culture* (pp. 523–525). New York, NY: Routledge.

Embodied Cognition

This entry discusses the recent major turn in the study of cognition that has brought to center stage a view of cognition as fundamentally embodied; explains how it arose in reaction to the received view of cognition dominant so far, in which the body played at best a secondary role; shows its precursors and sources of inspiration; and underlines some of the recent evidence in its favor offered by cognitive neuroscience. What is important is that this new conception of cognition as embodied brings in its wake a wider concept of what cognition is, linking it inexorably to its environment, which is not just physical but also interpersonal and social. Brain, body, and environment thus form a multidimensional dynamic reality in which meanings, value, and culture are present.

Historical Background

The term *embodied cognition* did not come into common usage until the beginning of the 21st century. Throughout most of the 20th century, mainstream Anglo-American "analytic" philosophy gave almost no attention to embodiment, focusing instead primarily on conceptual and linguistic analysis and conditions of epistemic justification. The body was regarded as nothing more than a repository for the brain and a conduit for sense perception, and it was therefore deemed not to be constitutive of processes of understanding and reasoning. The dramatic reversal of this neglect of the body began partly in phenomenology, gained momentum in some parts of feminist philosophy, and came to fruition with the recent meteoric rise of the cognitive sciences over the past three decades. In particular, the biological and cognitive sciences have come to recognize how our embodiment gives rise to mind, thought, language, and other forms of symbolic human interaction. What was once the most peripheral of topics has now moved to center stage in virtually every discipline concerned with any aspect of human existence.

Fundamental Tenets

"Embodied cognition" does not define a single monolithic perspective, but most of those who align with this orientation would agree with two fundamental claims: (1) there is no mind or mental activity

(and hence no perception, feeling, thought, communication, valuing, or action) without a body and (2) the body plays an indispensable role both in *what* has meaning for us and in *how* we experience and make that meaning. In other words, the body does not just provide representational content for mental operations. Rather, it shapes the ways we conceptualize, reason, and communicate.

According to the first tenet, to say that cognition is embodied is to say that the body actively functioning within its environments is indispensably necessary for any kind of cognitive activity. The American pragmatist philosopher John Dewey argued that the locus of all experience, thought, and communication is a human organism in ongoing engagement with an environment that has physical (biological), interpersonal, and cultural dimensions intricately interwoven. Dewey coined the term *body-mind* to capture this continuous transaction between what we term the *physical* and *mental* aspects of our existence. In short, no body, never mind. Moreover, Dewey correctly insisted that we cannot adequately address the nature of "mind," experience, or thought without paying attention to the environments in which the mind emerges and operates. In short, no world, never mind. Today, the preferred term for this ongoing embodied experiential process is *enaction,* which is meant to capture the dynamic, active, purposive engagement of an organism with its ever-changing environments.

Embodied cognition views emerged as a criticism of, and a major alternative to, theories in the 1960s and 1970s that regarded the mind as a unified set of computational programs that take input and generate output. These early computational ("functionalist") theories of mind meshed nicely with the then-dominant Chomskyan generative linguistics paradigm, which focused on an allegedly innate mechanism for generating formal structures (syntax) for possible natural languages. Chomsky claimed that syntax is independent from anything having to do with our bodies, and so he incisively described his project as a "Cartesian linguistics." Functionalist theories tend to ignore the body, claiming only that some embodied "wetware" is needed, but only to run the programs that constitute mind, thought, and language processing.

By the 1970s, it was becoming clear that these functionalist approaches were incompatible with the newly developing sciences of mind that began to show how our bodily capacities and actions provide the basis for human meaning, reasoning, and symbolic interaction. According to this new body of empirical research, our bodies are the primary source of meaning, values, and the very patterns and qualities of our so-called mental operations.

This developing research generates the second tenet above, which is that the body in large measure determines both *what* we can experience, conceptualize, reason about, and communicate and *how* we perform these cognitive activities. The body plays an indispensable role in the possibility and nature of all aspects of cognition.

Sources

Evidence for the embodiment of mind, thought, and language comes primarily from three sources: (1) phenomenological descriptions of our bodily situatedness in the world, (2) studies of the conceptual systems that underlie language and other forms of symbolic interaction, and (3) the cognitive science and neuroscience of the body and brain. Taken together, these bodies of research make a compelling case for the central role of the body in every aspect of human cognition.

Phenomenology

Although the forefathers of phenomenology—Edmund Husserl, Martin Heidegger, and other existential phenomenologists—recognized the body's role in shaping what they called our *lifeworld,* it was the French phenomenologist Maurice Merleau-Ponty who did the most to remind us that our body is never merely a physical object in space but rather a "lived" or "phenomenal" body that provides the very possibility of our inhabiting a meaningful world and is the source of our meaning and thought. All of the dichotomous distinctions we make, such as mind/body, subject/object, cognition/emotion, are abstractions from the bodily processes out of which our world emerges. They mark real aspects of our experience, but they have no substantial independent existence. Merleau-Ponty, thus, put the lived body center stage by recognizing it as the corporeal and interpersonal processes that give rise to all the meaning and significance of which we are capable, including gesture, language, science, and all the arts. A good example of this type of approach is Maxine Sheets-Johnstone's exploration of the role of bodily

movement in our ability to make sense of our world. She shows how this "primal animateness" both situates us and generates the qualities, properties, and relations that define how we dwell meaningfully in our world.

Empirical Studies of Embodied Conceptual Systems and Language

If "body" and "mind" are not distinct, separable *things* but are instead dimensions, aspects, or patterns abstracted from the primordial give-and-take of an organism and its environment, then what we call conceptualization is a fundamentally bodily process. Lawrence Barsalou's exploration of perceptual symbol systems reveals how our concepts and reasoning are inherently rooted in perceptual and motor activities. He emphasizes the key role of sensory-motor simulation in our ability to experience and make meaning. George Lakoff and Mark Johnson have identified a number of basic body-based "image schemas"—such as up/down (verticality), containment, source–path–goal, iteration, front/back, scalar intensity, and so forth—that structure our perception, action, and spatial relations concepts (and their linguistic expressions). For instance, the English word *in* activates a container schema with its own distinctive logic of in–out relations, based on our bodily experience of how physical containers (e.g., boxes, cups, bowls) give rise to a bodily logic of containment (e.g., *if Container A is in Container B, and Container B is in Container C, then A is in C*—known as the Transitivity relation in logic).

The bodily grounding of concepts for concrete objects, spatial relations, and actions is perhaps not such a surprising hypothesis, since our sensory-motor experience is so basic to our ability to understand, act within, and transform aspects of our physical environment. What is more striking is the role of embodiment in creating and understanding abstract concepts (e.g., mind, thought, causation, values, institutions, mathematics, and logic). One of the most productive strategies for explaining abstraction has come from Conceptual Metaphor Theory, which argues that virtually all our abstract concepts are defined by metaphorical mappings of structure from a typically body-based source domain onto a target domain, so that the image-schematic spatial or corporeal logic of the source domain is used to reason about the target. For instance, the metaphor "Knowing is seeing" conceptualizes acts of knowing as body-based visual processes, as in "I *see* what you mean" and "Could you *shed some light* on relativity theory?"

Cognitive Neuroscience

The development of sophisticated neuroimaging technologies has recently made it possible to supplement traditional phenomenological description, conceptual analysis, and cognitive linguistic research by revealing some of the bodily systems underlying these conceptual systems and thought processes. By means of lesion studies and brain-imaging experiments, we are beginning to understand some of the general conditions for certain types of cognitive activity, such as perception, conception, feeling, emotion, reasoning, communication, and willing. One of the more striking recent findings is about the key role played by sensory-motor processes, emotions, and feelings in reasoning. In contrast to disembodied views of cognition, neuroscientists like Antonio Damasio and Gerald Edelman present extensive research showing that what we traditionally have thought of as purely "mental" cognition actually recruits neural clusters responsible primarily for perception and action. It appears that a functional emotional apparatus is necessary for practical and social reasoning.

Recent exciting developments in cognitive neuroscience have also made it possible to give the basics of an embodied cognition approach to language and symbolic interaction. Based on the joint tenets that (a) thought is structured neural activity and (b) language is inextricable from thought and experience, Jerome Feldman and several others have explored what are known as structured or constrained connectionist neural models for various aspects of syntax, semantics, and pragmatics of natural languages. George Lakoff and his coauthors have provided the basics of similar neural models for certain types of concrete concepts, image schemas, and conceptual metaphors. The cognitive neuroscience of language and symbolic interaction is still in its infancy, but it is making major strides in uncovering the neural bases of our capacities for language comprehension and production. Most promisingly, it is beginning to offer experimental methods for determining whether sensory and motor areas are actually activated in the processing of various kinds of concepts and in our reasoning about abstract notions.

In only a few decades of research, we have thus moved from an almost total neglect of embodiment to a deep appreciation of the role of the body in mind, thought, and language. First, what we call "mind" and "mental activity" are intrinsically embodied. Second, as Gerald Edelman has insisted, it is now evident that neither is the brain completely modular nor is all cognition processed in a global, whole-brain fashion. There exist both modularity and complex reciprocal parallel interactions of multiple brain regions for most of our cognition. Third, thought and feeling are intimately intertwined, so much so that an intact emotional system is a requisite for all forms of practical and social reasoning. The traditional rigid and exclusive alignment of the mind with thought (conceptualization and reasoning) and the body with feeling and emotion is a nonstarter from a neural perspective. The body shapes all aspects of our higher cognitive activities. Fourth, the mind is not the brain. As William Bechtel has argued, although there is a tendency toward reductionist explanation in cognitive science, reductionist accounts of cognition often recognize the necessity of multiple irreducible explanatory levels that require multiple methods of inquiry and theorizing. None of this reduces the mind to neural activation alone.

We, thus, return to our earlier insistence that the locus of experience and cognition is not just a brain, not just a body, and not just a system of cultural practices but rather a functioning brain in a functioning body engaged with environments—physical, interpersonal, and cultural. The "body" of embodied cognition is a complex, multidimensional dynamic reality that gives us meaning, understanding, thought, and value.

Mark Johnson

See also Being-in-the-World; Classical Computationalism, Connectionism, and Computational Neuroscience; Cognitive Phenomenology; Cognitive Sciences; Distributed Cognition and Extended-Mind Theory; Grounded Cognition and Social Interaction; Life-World; Mind–Body Relation; Neural Hermeneutics; Situated Cognition; Social Perception

Further Readings

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience.* New York, NY: Psychology Press.

Damasio, A. (2010). *Self comes to mind: Constructing the conscious brain.* New York, NY: Pantheon.

Dewey, J. (1981). Experience and nature. In J. Boydston (Ed.), *John Dewey: The later works, 1925–1953* (Vol. 1, pp. 361–364). Carbondale: Southern Illinois University Press. (Original work published 1925)

Edelman, G. (2005). *Wider than the sky: The phenomenal gift of consciousness.* New Haven, CT: Yale University Press.

Feldman, J. (2006). *From molecule to metaphor: A neural theory of language.* Cambridge: MIT Press.

Johnson, M. (2007). *The meaning of the body: Aesthetics of human understanding.* Chicago, IL: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought.* New York, NY: Basic Books.

Merleau-Ponty, M. (1962). *Phenomenology of perception* (C. Smith, Trans.). London, England: Routledge.

Tucker, D. (2007). *Mind from body: Experience from neural structure.* Oxford, England: Oxford University Press.

Emergence

An emergent phenomenon is one that is novel, develops from other (usually more basic) phenomena, has some degree of autonomy from those phenomena, and often has a holistic aspect.

Because social phenomena are often significantly different from psychological and biological phenomena, can frequently be given a functional characterization, and often appear to possess holistic features, the first, third, and fourth of these properties tend to be present in social systems. The constructive task is then to provide an account of how the second feature is satisfied. In earlier times, emergent phenomena were viewed with suspicion, in part because of a bias toward reductionist methods, including methodological individualism, and in part because early writers on the topic restricted themselves to problematic cases such as consciousness and vital forces. The suspicion has lessened as a result of an increased understanding of what the concept of emergence entails and with the realization that emergence may be quite common, allowing more transparent examples.

Taxonomy of Emergence

What counts as an example of emergence is, nevertheless, currently a matter of significant dispute. One can divide positions on emergence into two broad kinds.

Emergence Resulting From Lack of Knowledge of a System

From the perspective of the first kind, emergence results from our lack of knowledge of a system. With systems of a sufficient degree of complexity, features can appear as a result of interactions between the components of the system, and those features will strike the observer as "novel" and "irreducible" because they lie far beyond our capacity to predict their appearance. For example, it is quite easy to construct computer simulations that effectively mimic the emergence of conventional standards in society, such as the almost universal adoption of the relatively inefficient QWERTY keyboard as a standard for typewriters. In the actual historical situation, the number and complexity of the choices made by typists, typewriter manufacturers, and businesses precluded both prediction of the emergence of this standard and its explanation via the decomposition strategies that are often successful in giving us understanding in less complex systems. Some versions of this unpredictability approach suggest that it is impossible in principle to predict emergent features, although because predictability is relative to the theory used for prediction and an unpredictability claim can be trivially undermined by adding the sentence to be predicted, such claims are rarely both precise and convincing. One reasonable interpretation of what is intended, however, is that the novel state of affairs could have been known only by waiting until it first appeared, even in the simulation model. These approaches using unpredictability fall into the category of what is often called *weak emergence*—weak because there is no commitment to downward causation, only to the existence of certain stable patterns in society.

What is, in contrast, often called *strong emergence* has a robustly ontological commitment; something genuinely new has emerged in the social world, beyond what was in the elements from which the whole was constructed. It is strong emergence that has traditionally been the source of disputes in the social sciences. Those who subscribe to some form of individualism—the position that only individuals and their properties exist—hold that not only do individuals preexist social arrangements, both temporally and ontologically, but that there is nothing over and above individuals, their properties, and the relations between individuals. Strong emergentists deny this and argue that there is more to the systems than individuals and their properties but tend to differ on what more there is. Some allow that there are social properties in addition to individual properties; others allow that higher-order entities such as corporations and universities exist in addition to their members; some others allow that individuals play no essential role in social processes and that all identifiably social causation occurs at the social level; and still others hold that there are emergent social laws that govern the development of social units or individuals. Within the individualist camp, the status of social relations is the subject of discussion. For some, such relations can be completely reduced to, or supervene on, intrinsic properties; in the latter case, once the intrinsic properties have been fixed, the social relations are automatically determined. For others, these social relations are sui generis and constrain the individuals they relate, often by influencing the agents' beliefs.

It is widely believed that the supervenience position precludes downward causation and hence strong emergence, because any influence from downward causation would be redundant given the causal processes at the level of individuals. But there is less reason to think that the realm of intentional action by individuals is causally closed in the way the physical realm might be closed, and so this application of the causal exclusion argument is not completely convincing.

Many anti-individualist positions can sensibly be called emergentist positions, and even individualist positions allow for weakly emergent features. Despite this, one must be aware that much talk of emergence in the social science literature is a *façon de parler* and does not reflect the existence of emergence in the strong sense.

It is commonly held among weak emergentists that emergence occurs when a threshold of complexity is crossed in a system. Investigations of this sort of emergence have been aided by the development of agent-based models. The constituent elements of these models are individuals, but in virtue of allowing multiple iterations of usually pairwise and often

nonlinear interactions between large numbers of agents, the emergence of novel structures or equilibria is sometimes observed. A well-known early example of this type of model was Thomas Schelling's toy model of social segregation, which demonstrated that small preferences about the type of one's neighbor can lead to structures of segregation in the society as a whole. These models have great potential in sociology, anthropology, and microeconomics, although they have not been welcomed by all, especially those who work in traditional economics. Such simulations support one version of weak emergence in which the only way to know how the system will evolve is to let the system run its course or to simulate its micro-states. This computational incompressibility gives some precision to unpredictability, but it is important to maintain a separation between emergent phenomena occurring in computer simulations of societies and the related phenomena occurring in the societies themselves. One consequence of our inability to predict the dynamics of complex systems on the basis of micro-information alone is the introduction of a macro-level vocabulary that is conceptually different from the micro-level descriptions and that facilitates prediction and description by virtue of erasing many of the details of the micro-states. This is why the vocabularies of microeconomics and of macroeconomics differ, as do the vocabularies of cognitive psychology and sociology and those of sociobiology and anthropology.

A common position holds that emergence entails the impossibility of reduction, although this incompatibility is occasionally resisted. This tradition requires that a hierarchy of levels or domains be identified for one level or domain to be reduced to or emerge from another. The usual ordering of levels based on composition, in which a necessary condition for one entity to exist at a higher level than another is that the latter be a constituent of the former, breaks down for the social sciences. Prima facie, one could argue that a particular type of economic system emerged from a set of cultural norms in a society, but neither economic processes nor cultural norms are constituents of one another. A possible source of confusion also needs to be addressed. It is sometimes claimed that *B*'s being explained in terms of *A* entails that *B* has been reduced to *A*. In certain cases, such as theoretical explanation, this makes sense, but in other cases, such as ontological explanations, this is straightforwardly false, as can be seen in the case when *A* is the cause of *B,* and the claim relating explanation and reduction seems to presuppose the presence of inter-level explanations.

Synchronic Versus Diachronic Emergence

A second broad division in emergence is that between synchronic and diachronic emergence. Although the majority of the philosophical literature on emergence addresses the synchronic case, the role of history in social institutions suggests the need to examine diachronic emergence more closely. Path dependence is a common phenomenon in the development of societies and social institutions, and synchronic emergence is unable to address this feature. It is the ability to computationally model the dynamics of social systems that has led to an increased acceptance of the existence of dynamically emergent social features.

Social Facts

The taxonomy given above makes no reference to rationality, intentionality, cultural norms, or other special features of human agency, whereas the existence of some emergent features of social groups seems to depend on the specifically social and normative aspects of social systems. The most prominent of these types of claims is Émile Durkheim's view that social facts exist, either as concrete institutions or as social norms, but social cooperation, group intentionality, and collective agency have been the focus of research in recent years. A core topic in this area is the emergence of cooperative behavior in societies, in particular the problem of how self-interested individuals can maintain a stable, cohesive, and efficient society. In this and other cases, normative criteria, especially those related to rationality, must frequently be taken into account. Some approaches in this area are reductive and neither require nor entail emergent features, but many are committed, via an irreducible appeal to first-person plural perspectives—"We endorse the candidate" can be true even if some individuals in the group voted against that candidate—to, at least, representations of collective social phenomena. Social interaction often involves representations, hence it presents the difficult philosophical problem of intentionality in ways that examples of physical emergence do not. To this end, John Searle has argued that collective intentionality resulting in linguistic declarations can

lead to social facts such as the establishment of a corporation, which may or may not be followed by the existence of an entity, the concrete corporation.

Although much of this literature concerns group agency, and hence the question of whether the group acts rationally on at least some occasions, there is an ontological aspect to this issue, which is whether social agents emerge from the interactions of individuals and exist autonomously from those individuals. For example, there is a long tradition within U.S. law of treating corporations as individuals. Whether or not this is legally or morally correct, one reason to think that some organizations are autonomous is the fact that group agents, such as corporations and universities, can retain their identities and ability to act despite complete replacement of the individuals who work in them. Mechanisms of self-organization can help us understand how certain types of groups can act cooperatively, and in Niklaus Luhmann's work, it is this combination of self-organization (*autopoiesis*) and the independence of features such as economic patterns from the individuals who instantiate them that is characteristic of social systems.

Despite the increased availability of tools to represent emergent features in social systems, it is important to maintain reductive individualism as the default position and to postulate emergent phenomena only when there is no convincing alternative position available. Unless one adopts this methodological position, introducing emergent entities becomes an escape route that is too easily available to avoid the difficult work of constructing a plausible individualist explanation.

Paul Humphreys

See also Agent-Based Modeling and Simulation in the Social Sciences; Collective Agents; Complexity and the Social Sciences; Emergence and Social Collectivism; Holism, in the Social Sciences; Individualism, Methodological; Reduction and the Unity of Science; Reductionism in the Social Sciences; Supervenience

Further Readings

Bedau, M., & Humphreys, P. (Eds.). (2008). *Emergence: Contemporary readings in philosophy and science.* Cambridge: MIT Press.

List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents.* New York, NY: Oxford University Press.

Sawyer, R. K. (2005). *Social emergence: Societies as complex systems.* Cambridge, England: Cambridge University Press.

Searle, J. (1995). *The construction of social reality.* New York, NY: Free Press.

Tesfatsion, L. (2002). Agent-based computational economics: Growing economies from the bottom up. *Artificial Life, 8,* 55–82.

Tuomela, R. (2002). *The philosophy of social practices: A collective acceptance view.* Cambridge, England: Cambridge University Press.

Emergence and Social Collectivism

Emergence refers to the processes whereby the global behavior of a system results from the actions and interactions of agents. There is no central controller or plan; a higher-level order emerges from the interaction of the individual components. Such systems are self-organizing, with control distributed throughout the system. Emergent systems are often complex in that they manifest order at the global-system level, which is difficult to explain by analyzing the individual components of the system in isolation. In this sense, sociological theories of emergence have been used to argue in support of social collectivism: the position that social collectives cannot be understood by reduction to properties of the participating individuals—a contrasting stance known as *methodological individualism*.

This entry summarizes emergence theories, and their connections to complexity theory and systems theory, and shows their relevance for the study of social systems, in particular social entities that are irreducible social collectivities.

Emergence, Complexity, and Systems Theory

Emergence is generally associated with complex systems—systems that reside between simplicity and randomness. When the laws governing a system are relatively simple, the system's behavior is easy to understand, explain, and predict. These systems are often most easily explained by reduction to an understanding of their components and their interactions. At the other extreme, some systems seem to behave randomly. There may be laws governing

their behavior, but the system is highly nonlinear, such that small variations in the state of the system at one time could result in very large changes to later states of the system. Such systems are often said to be *chaotic*. Complex systems are somewhere in between these two extremes: The system is not easy to explain by reduction to an analysis of its components, but it is not so chaotic that understanding it is completely impossible.

Examples of emergence include traffic jams, the colonies of social insects, and bird flocks. To illustrate, the V shape of the bird flock does not result from one bird being selected as the leader and the other birds lining up behind the leader. Instead, each bird's behavior is based on its position relative to the birds nearby. The V shape is not planned or centrally determined in top-down fashion. Rather, it emerges out of simple pair-interaction rules, that is, from the bottom up. The bird flock demonstrates one of the most striking features of emergent phenomena: Higher-level regularities are often the result of quite simple rules and local interactions at the lower level.

Mental states, such as conscious awareness, memories, or intentions, are often said to emerge from the biological brain. Complex systems that manifest emergence tend to have a large number of units, with each unit connected to a moderate number of other units, and frequent, repeated interactions among the connected units, which occur simultaneously throughout the system. Continuing our example of the mind–brain relation, the brain is a complex system composed of more than 100 billion neurons, with each connected to between 1,000 and 10,000 other neurons at synapses. Complex physical and biological systems tend to have relatively simple interactions between components. For example, neurons communicate by changing the rate of firing across a synapse. In contrast, the units in complex social systems are individuals who communicate using the full richness of natural language.

Emergence and Society

In the social sciences, a comparable example of an emergent phenomenon is language shift. Historians of language have documented that languages have changed frequently throughout history, with vocabulary and even grammar changing over the centuries. Yet until the rise of the modern nation-state, such changes were not consciously selected by any official body, nor were they imposed by force on a population. Rather, language shift is an emergent phenomenon, arising out of the nearly infinite number of everyday conversations in small groups scattered throughout the society. In this social system, successive conversations among speakers result in the emergence over time of a collective social fact: language as a property of a social group. The study of social emergence requires a focus on multiple levels of analysis—individuals, interactions, and groups—and a dynamic focus on how social group phenomena emerge from communication processes among individual members.

Societies have often been compared with other complex systems that manifest emergence. Just after World War II, Talcott Parsons's influential structural functional theory was inspired by cybernetics, a field centrally concerned with developing models of the computational and communication technologies that were emerging in the postwar period. In the 1960s and 1970s, general systems theory continued in this interdisciplinary fashion. It was grounded in the premise that complex systems at all levels of analysis—from the smallest unicellular organisms up to modern industrial societies—could be understood using the same set of theories and methodologies.

Common to all of these metaphors is the basic insight that societies gain their effectiveness and functions from a complex configuration of many people engaged in overlapping and interlocking patterns of relationships with one another. Some key questions raised by these "society as system" metaphors are as follows: What do these relations and configurations look like? Which systems are the most effective, and which are stable and long lasting? How could a stable complex system ever change and evolve, as societies often do? What is the role of the individual in the system? Such questions have long been central to sociology.

Theories of emergence have the potential to provide several new insights into these central sociological questions. Beginning in the mid-1990s, several scientific developments converged to create a qualitatively more advanced approach to emergence, and theories of emergence (often drawn from complexity theory) began to influence a wide range of disciplines from biology to economics. This influential new approach has begun to filter into

sociology. The study of emergence can provide new perspectives on important unresolved issues facing the social sciences—the relations between individuals and groups, the emergence of unintended effects from collective action, and the relation between the disciplines of economics and sociology.

Parsons's structural functional theory represented the first wave of systems theories in sociology, drawing on systems concepts from cybernetics to describe human societies as complex self-maintaining systems. The general systems theories of the 1960s and 1970s represented a second wave. General systems theories were always more successful at explaining natural systems than social systems. In spite of the universalist ambitions of such theorists, social scientists generally ignored them. In contrast, the latest work in complexity theory—the *third wave* of systems theory—is particularly well suited to sociological explanation because it focuses more squarely on emergence. Third-wave sociological systems theory grew out of developments in computer technology. In the 1990s, computer power advanced to the point where societies could be simulated using a distinct computational agent for every individual in the society, using a computational technique known as *multi-agent systems*. A multi-agent system contains hundreds or thousands of agents, each engaged in communication with at least some of the others. The researcher can use these simulations to create *artificial societies*. The researcher defines and implements a model of the individual agent, creates a communication language for the agents to interact, and then observes the overall macro-behavior of the system that emerges over time.

Whether or not a global-system property is emergent, and what this means both theoretically and methodologically, has been defined in many different ways. For example, in some accounts, system properties are said to be emergent when they are *unpredictable* even given complete knowledge of the lower-level description of the system—a complete knowledge of the state of each component and of their interactions. In other accounts, system properties are said to be emergent when they are *irreducible,* in any lawful and regular fashion, to properties of the system components. In yet other accounts, system properties are said to be emergent when they are *novel,* when they are not held by any of the components of the system. Philosophers of science began debating such properties early in the 20th century. Social scientists have applied widely different definitions of emergence, resulting in some conceptual confusion.

Emergence Versus Reductionism: Social Collectivism

Sociological theorists and philosophers of science have argued that the emergent higher level may have autonomous laws and properties that cannot be easily reduced to lower-level, more basic sciences. Thus, the paradigm of complexity is often opposed to the paradigm of reductionism. For example, cognitive scientists generally agree that mental properties may not be easily reduced to neurobiological properties, because of the complex dynamical nature of the brain. In an analogous fashion, several sociological theorists have used complexity theory to argue against attempts to explain societies in terms of individuals, a reductionist approach known as *methodological individualism*. Because many socially emergent phenomena are difficult to explain in terms of the system's components and their interactions, these theorists have claimed that emergentist thinking supports social collectivism or holism and that individualist approaches will have limited success as a potential explanation for many social phenomena.

For example, due to complexity and emergence, there may be potential limitations of individualist methodologies such as neoclassical microeconomics and evolutionary psychology. Emergence and social collectivism suggest that both psychology and microeconomics are likely to be severely limited in their ability to explain human behavior in groups. As currently conceived, psychology is the study of system-independent properties of individuals (e.g., variables, traits, mental models, cognitive capacities). Microeconomics is the study of how collective phenomena emerge from aggregations of individual preferences and actions. Both are individualist in that they reject explanations that propose that group properties could lawfully influence individual action. Many contemporary paradigms are based on such reductionist assumptions—evolutionary psychology, cognitive neuroscience, behavioral genetics, and social cognition. Yet an emergentist perspective suggests that many social systems may not be explainable in terms of individuals and that neither

psychology nor microeconomics can fully explain the socially contextualized nature of human behavior.

Because societies are complex systems, individualists cannot assume that a given social system will be reducible to explanations in terms of individuals. Inversely, social collectivists cannot assume that a given social system will not be so reducible. Whether or not a social system can be understood solely in terms of its component individuals and their interactions is an empirical question, to be resolved anew with respect to each social system. Theories of emergence show why some social properties cannot be explained in terms of individuals. Thus, one cannot assume that methodological individualism can exhaustively explain human behavior in social groups. However, not all social systems are irreducibly complex, and some social properties can be explained by identifying their processes of emergence from individuals in interaction. Complexity approaches can help determine which approach will be the most appropriate for which social system.

Conclusion

Studies of social groups must be fundamentally interdisciplinary, because a focus on emergence requires a simultaneous consideration of multiple levels of analysis: the individual, the communication language, and the group. A complete explanation of the most complex social systems may require interdisciplinary teams composed of psychologists, sociologists, communication scholars, and economists.

R. Keith Sawyer

See also Analytical Sociology and Social Mechanisms; Complexity and the Social Sciences; Emergence; Feedback Mechanisms and Self-Regulatory Processes in the Social Sciences; Holism, in the Social Sciences; Individualism, Methodological; Mechanism and Mechanismic Explanation; Reductionism in the Social Sciences; Supervenience; Systems Theory

Further Readings

Archer, M. S. (1995). *Realist social theory: The morphogenetic approach*. New York, NY: Cambridge University Press.

Johnson, S. (2001). *Emergence: The connected lives of ants, brains, cities, and software*. New York, NY: Scribner.

Sawyer, R. K. (2001). Emergence in sociology: Contemporary philosophy of mind and some implications for sociological theory. *American Journal of Sociology, 107,* 551–585.

Sawyer, R. K. (2006). *Social emergence: Societies as complex systems*. New York, NY: Cambridge University Press.

Thrift, N. (1999). The place of complexity. *Theory, Culture, and Society, 16,* 31–69.

Waldrop, M. M. (1992). *Complexity: The emerging science at the edge of order and chaos*. New York, NY: Simon & Schuster.

Emotions

Emotions have been one of the subject matters of philosophy from ancient times to the present. This entry presents two of the main ways of analyzing the essential features of emotional states and concludes by delineating how emotions relate to morality.

Emotions are probably the most complex mental phenomena, as they involve all types of mental entities and states that belong to various ontological levels. In order to capture the complexities and subtleties of emotions, a few complementary methods may be used. One major way is to describe a typical emotion; another is to attempt to define the essence of emotions.

What Is a "Typical" Emotion?

Typical emotions are generated by perceived significant changes, and their focus of concern is personal and comparative. Typical emotional characteristics are instability, great intensity, partiality, and brief duration. Basic components are cognition, evaluation, motivation, and feelings.

Emotions typically occur when we perceive positive or negative significant changes in our personal situation or in the situation of those related to us. Like burglar alarms going off when an intruder appears, emotions signal that something needs attention.

Emotions occur when a change is evaluated as relevant to our personal concerns. Emotions serve to monitor and safeguard our personal concerns; they give the eliciting event its significance. Emotional meaning is mainly comparative. The comparison underlying emotional significance encompasses the mental construction of the availability of an alternative situation. The more available the

alternative—that is, the closer the imagined alternative is to reality—the more intense the emotion.

Instability, great intensity, a partial perspective, and relative brevity can be considered as the basic characteristics of typical emotions. This characterization refers to "hot emotions," which are the typical intense emotions. The more moderate emotions lack some of the characteristics associated with typical emotions.

In light of the crucial role that changes play in generating emotions, instability of the mental (as well as the physiological) system is a basic characteristic of emotions. Emotions indicate a transition in which the preceding context has changed but no new context has yet stabilized. Emotions are like storms and fire—they are intense, occasional, and limited in duration.

Emotions are intense reactions. In emotions, the mental system has not yet adapted to the given change, and due to its significance, the change requires the mobilization of many resources. No wonder that emotions are associated with urgency and heat. A typical characteristic of emotions is their magnifying nature: Everything looms larger when we are emotional.

Emotions are partial in two basic senses: They are focused on a narrow target, such as one person or very few people, and they express a personal and interested perspective. Emotions direct and color our attention by addressing practical concerns from a personal perspective.

Typical emotions are relatively brief. The mobilization of all resources to focus on one event cannot last forever. A system cannot be unstable for a long period and still function normally; it could explode due to the continuous increase in emotional intensity.

In addition to these typical characteristics, we can also divide emotions into four basic components: cognition, evaluation, motivation, and feeling. Cognition contains descriptive information about the object; evaluation assigns value to the information included in the cognitive component and, consequently, in many cases generates great motivation and desire to act in a certain manner toward the emotional object. The feeling component expresses the subject's state.

Emotional intensity is determined by several variables that can be divided into two major groups: one referring to the perceived impact of the event eliciting the emotional state and the other to the background circumstances of the agents involved in the emotional state. The major variables constituting the event's impact are the strength, reality, and relevance of the event; the major variables constituting the agent's background circumstances are accountability, readiness, and deservingness.

Kindred Types of Affective States

Together with the emotions, the affective realm includes other phenomena such as sentiments, moods, affective traits, and affective disorders. Emotions and sentiments have a specific object, whereas the object of moods, affective disorders, and affective traits is general and diffuse. Emotions and moods are essentially occurrent states; sentiments and affective traits are dispositional in nature. These differences are expressed in temporal differences. Emotions and moods are relatively short, whereas sentiments and affective traits last for a longer period.

The Common Essence of Emotions

In addition to explaining emotions by describing their typical features, emotions can be explained by defining the essence common to all those features. This method is problematic because emotions are so complex that it is difficult to find such a single definition.

One such possible definition is that emotions are a general mode (or style) of the mental system. Such a general mode is a complex arrangement that functions over time. It is a dynamic, structured experience. The kinds of elements involved in a certain mode and the particular arrangement of these elements constitute the uniqueness of each mode. Other possible modes are the perceptual, imaginative, and intellectual modes. The emotional mode is more complex, comprehensive, and dynamic than the other mental modes.

The relationship between emotional and intellectual modes has been at the center of many disputes. There is a long tradition that criticizes emotions as being irrational and nonfunctional. It seems that this tradition is fundamentally wrong. Emotions are rational in the normative sense of being an appropriate response in the given circumstances. Although emotions are not functional in all circumstances, they are tremendously important when facing urgent situations involving a significant change. In these situations, emotions might be the optimal response.

Three major functions of emotions are (1) an initial indication of the proper direction in which to respond, (2) quick mobilization of resources, and (3) social communication. Emotional excess might be harmful, but so are all types of excess. Thus, it is advisable neither to suppress our emotions nor to have an excess of them; ideally, we should strive for emotional balance.

Emotional intelligence is the optimal integration of the emotional and intellectual systems; it consists of recognizing and regulating emotions in an optimal manner. In light of the differences between the two systems, we can speak of "emotional reasoning" as different from intellectual reasoning.

Emotions and Morality

The role of emotions in the moral domain has been frequently disputed. The partial nature of emotions seems to contradict the more general and egalitarian nature of basic moral evaluations. Emotions are nevertheless morally valuable. They are especially so in our relationships with those near and dear to us. In such circumstances, which constitute the bulk of our daily life, discriminatory emotional attitudes are not only possible but morally commendable as well. Particular emotional attention to the specific needs of those close to us is of crucial moral importance. Emotional attitudes are also a moral barrier against many crimes. The crucial role of emotions in moral life does not imply their exclusivity; the intellectual capacity is important as well.

Aaron Ben-Ze'ev

See also Affective Intelligence in the Social Sciences; Collective Emotions; Emotions in Economic Behavior; Game-Theoretic Modeling; Love, in Social Theory; Love, Philosophy of

Further Readings

Ben-Ze'ev, A. (2000). *The subtlety of emotions*. Cambridge: MIT Press.

Frijda, N. H. (2006). *The laws of emotion*. London, England: Psychology Press.

Goldie, P. (Ed.). (2010). *The Oxford handbook of philosophy of emotion*. Oxford, England: Oxford University Press.

Nussbaum, M. C. (2001). *Upheavals of thought: A theory of the emotions*. Cambridge, England: Cambridge University Press.

Oatley, K., Keltner, D., & Jenkins, J. M. (2006). *Understanding emotions*. Malden, MA: Blackwell.

Emotions in Economic Behavior

Emotions can influence economically relevant decisions in at least two ways. First, emotions expected to occur as the result of a particular course of action (e.g., happiness that might result from taking a vacation or regret that might result from not taking it) can be components of that action's desirability or "utility." Second, emotions experienced at the moment of choice may influence expected utility or override deliberate consideration of expected utility. These "immediate emotions" can either be integral to the decision at hand (e.g., the anxiety experienced when deciding whether to make an indulgent purchase) or incidental to the decision (e.g., lingering anxiety about an upcoming medical exam). Economists have made good progress toward understanding the influence of "expected emotions" on decision making but have only recently explored the role of immediate emotions.

Insights From Behavioral Economics

For most of the past century, economists largely ignored emotions. Economists typically assumed that decisions reveal stable preferences. Because choice presumably conveys such rich information, expected emotions, which may merely be one component of stable preferences, required little attention. Emotions experienced at the moment of choice were of even less interest, since decision makers should be forward looking, forming their decisions purely on the basis of potential outcomes. Other properties of emotions such as their fleeting nature and the difficulty involved in measuring them precisely likely contributed to their absence from economic analyses.

This situation began to change in the 1980s with the emergence of behavioral economics, a subfield of economics that incorporates insights from psychology and neighboring disciplines to increase the descriptive accuracy and predictive power of economic theory. Findings from behavioral decision research, a subfield of psychology, increasingly suggested that

decision makers rarely act on preferences that are stable and well formed but rather on preferences that are constructed on the spot. Contextual factors that influence the construction of preferences (e.g., whether outcomes are framed as gains or losses) generated much interest among behavioral economists. Emotions were less central to behavioral economists but did begin to receive attention.

Behavioral economists appealed to emotions to help explain phenomena that were anomalous from the standard economic perspective. Consider, for example, typical behavior in the "ultimatum game." In the game, there is an amount of money to be divided between a proposer and a responder. The proposer proposes a division, and the responder can either accept the proposal or reject it, leaving both players with nothing. If players are strictly self-interested (the standard economic assumption), the responder should accept any positive amount of money, and the proposer should therefore offer the responder the smallest positive amount of money possible. In reality, proposers often offer to split the money equally, and responders often reject lopsided offers that would leave them with far less than the proposer. The results suggest that anger over being treated unfairly can swamp self-interest (and proposers largely expect this). Indeed, Sally Blount found that responders were significantly more likely to accept lopsided proposals when those proposals were generated by chance than when those proposals were actually made by human proposers, suggesting that anger played a significant role in the high rejection rate of intentionally lopsided proposals.

Other research demonstrated that incidental emotions, unrelated to the task at hand, could predictably influence economically relevant decisions. For example, David Hirshleifer and Tyler Shumway found that the amount of morning sunshine correlated positively with daily stock returns, possibly because investors misattributed their good mood to positive economic prospects. Laboratory research by Jennifer Lerner and colleagues also demonstrated that discrete incidental emotions can influence economic valuations (e.g., willingness to pay), further challenging standard economic theory.

Enter Neuroeconomics

Interest among economists in the role of immediate emotions has increased markedly since the advent of neuroeconomics, a subfield of behavioral economics that examines the neural processes that underlie economic decision making. Early foundational work in this area by Antoine Bechara, Antonio Damasio, and colleagues focused on the role of *somatic markers,* feelings that encode the consequences of alternative courses of action, in guiding decision making. They demonstrated that people with damage to the ventromedial prefrontal cortex, a brain region that plays a critical role in this affective encoding process, were unable to execute what they cognitively realized was the optimal strategy in an economic game, thus costing themselves money. While the extent to which emotional deficits actually harm decision making depends on the situation, this work demonstrates that immediate integral emotions can play a central role in decision making.

Interest in emotions further accelerated as neuroeconomists began to investigate the neural underpinnings of economic decision making with functional magnetic resonance imaging. Economists, chronically skeptical of the "cheap talk" possible when experimental participants are asked to report their feelings or why they behaved in a particular way, are increasingly willing to listen to the brain rather than the person. There are limitations to neuroeconomic data, however (e.g., they are largely correlational; inferring an emotion from activation in a particular region is difficult when that region is not selectively activated by that emotion), and the subfield is certainly not without its critics within economics, who argue that economics should remain focused on predicting behavior rather than the processes underlying behavior (e.g., neural activation). Nevertheless, a spate of models inspired by neuroeconomic research, attempting to characterize the interactions between affect and deliberation, has recently emerged in behavioral economics.

Open Questions

Ongoing debates center on whether there are separate emotional and deliberative neural systems that interact to influence decision making, whether emotions help or hurt decision makers, and whether societal problems are more attributable to intense or insufficient emotions.

Scott Rick

See also Affective Intelligence in the Social Sciences; Decision Theory; Emotions; Heterodox Economics; Neuroeconomics; Philosophy of Economics, History of; Social Choice Theory

Further Readings

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*(1–3), 7–15.

Caplin, A., & Leahy, J. (2001). Psychological expected utility theory and anticipatory feelings. *Quarterly Journal of Economics, 116*(1), 55–79.

Hirshleifer, D., & Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *Journal of Finance, 58*(3), 1009–1032.

Lerner, J. S., Small, D. A., & Loewenstein, G. (2004). Heart strings and purse strings: Carryover effects of emotions on economic decisions. *Psychological Science, 15*(5), 337–341.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes, 65*(3), 272–292.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review: Papers and Proceedings, 90,* 426–432.

Mellers, B., Schwartz, A., & Ritov, I. (1999). Emotion-based choice. *Journal of Experimental Psychology: General, 128*(3), 332–345.

Rick, S., & Loewenstein, G. (2008). Intangibility in intertemporal choice. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1511), 3813–3824.

Empathy

The British psychologist Edward Titchener introduced the term *empathy* in 1909 as an English translation of the German word *einfühlung* ("feeling into"). *Einfühlung* is rooted in philosophical aesthetics. It was used by German philosophers toward the end of the 19th century to describe our ability to imaginatively "feel into" works of art and nature. However, it was the German philosopher Theodor Lipps who broadened the term to encompass our experience of other people. Lipps transformed empathy from a predominantly aesthetic concept into a concept at the center of philosophical and psychological analyses of sociality. This emphasis on the social-scientific significance of empathy continues to inform current discussions.

This entry looks at empathy within the context of social cognition. It considers empathy in relation to the philosophical problem of other minds, the mechanisms of social cognition, and the relation between empathy and affectivity.

Empathy and the Epistemology of Other Minds

Empathy has been summoned to deal with the philosophical problem of other minds. This problem arises from the question "How do we know that other people have minds like ours?" That others *do* have minds like ours seems fairly certain. Yet one might think that we cannot see other minds or experience them via some other perceptual modality. The only mind we can directly experience is our own. Given this lack of experiential access to any mind but our own, how is knowledge of other minds possible?

This is an epistemological puzzle: the question of how we are justified in believing that others have minds like ours. One answer is that we rely on inference from analogy. This inference explains how we attribute minds to others. It begins with a Cartesian assumption: We enjoy direct, infallible access to our own mind; in contrast, our access to other minds is indirect and fallible. Moreover, we know that when we experience certain mental states (anger, sadness, etc.), those states characteristically cause certain patterns of behavior (frowning and fist shaking, weeping, etc.). Accordingly, when we observe this behavior in others, we infer the existence of the relevant mental state causally responsible for that type of behavior. We assume the other to be psychologically similar to us and infer that his behavior is animated by the same type of mental state(s) animating our own behavior.

Many have not found this answer satisfactory. Lipps is one such philosopher. In his 1907 article "Das Wissen von fremden Ichen" (The knowledge of other "I"s), Lipps argues that analogy cannot account for our basic openness to others. It is unclear, Lipps argues, how analogical inference can simultaneously allow us to think about another's mental states as both similar to our own and yet radically different, that is, as *that person's* mental states. Lipps argues that empathy can better explain this openness. For Lipps, this "instinct of empathy" consists of my imitating the gestures or expressions I see in others. When I see another's expression of anger,

say, I reproduce this anger—I experience the feeling of anger myself—but I then project this feeling onto the person who first evoked it. Empathy is therefore a process of simulation and projection. It is, Lipps argues, at the root of interpersonal understanding.

Phenomenologists such as Edmund Husserl, Max Scheler, and Edith Stein offer an alternative picture. They agree that analogical inference is not the way to understand our basic experience of other minds. Yet, although they differ in the details, they reject Lipps's "simulation plus projection" alternative. Instead, they argue that empathy is what Stein terms an "experiential act sui generis": a primitive, irreducible form of intentionality that, prior to inference or simulation, presents other human beings to me *as* "minded" *in* my experience of them. Empathy is a form of direct perception, not simulation. From the start, I perceive others differently than I do rocks, tables, or trees. I see them as a locus of unique thoughts, feelings, and intentions: a psycho-physical "expressive unity," as Scheler puts it. This is because the actions, gestures, and facial expressions of others present me with the experience of a concretely embodied mind. The phenomenological model of empathy thus rejects the Cartesian assumption that minds are entirely in the head, hidden from others. In perceiving expressive behavior, I see the mental life of others play out *in* that behavior. Perception of others in their concrete expressiveness is thus sufficient to justify our belief that others are likewise minded.

Empathy and the Psychology of Other Minds

There remains another puzzle. We might term this the psychological problem of other minds: the question concerning the mechanisms ultimately responsible for interpersonal understanding. Here, empathy is often framed as a kind of "mind reading," or the ability to detect and respond to the mental life of others and to interpret and predict their behavior.

This is an empirical question. It is distinct from the epistemological question of other minds. Independently of whether or not we are justified in believing in other minds, there remains the question of how we come to know what another person is thinking, feeling, or intending, or how to interpret his or her behavior. This issue is the focus of most current empathy research in the cognitive and social sciences. However, there is little agreement on the proposed candidate mechanisms enabling empathy. There is even disagreement over the nature of empathy itself.

We have already encountered some of the mechanisms said to be responsible for empathy: inference, simulation, and perception. *Theory Theorists* argue that we use our lay theories about how minds work ("folk psychology") to infer the existence of mental states in others and to interpret their emotional expressions and behavior. These lay theories are the basic mechanisms of empathy. They are sometimes said to emerge from innate 'mind reading' modules in the brain; others claim that they develop as we age and gain social experience. When we deploy our lay theories, we use inference from perceived behavior (including utterances) to hidden internal states. This theory-driven knowledge of another's internal state(s) is sometimes called "cognitive empathy" or "empathic accuracy."

Simulation theorists, on the other hand, argue that we use our own emotional and imaginative resources to put ourselves in others' mental shoes and read their internal states through observing their behavior. We use imagination as well as the knowledge of how we would feel in their situation to take their perspective and come to understand what they are thinking and feeling. This is sometimes called "projective empathy" or "perspective taking."

Other simulation theorists argue that a different sort of simulation is responsible for empathy: *mimicry*. I mimic another person's posture or expression —think of two friends speaking closely while leaning on a bar or a newborn imitating the facial expressions of her caregiver—and come to understand what this other person is thinking and feeling. This is an example of behavioral mimicry. It has been called "facial empathy," "imitation," or more commonly "motor mimicry." Other simulation theorists argue that the relevant mimicry occurs at the neural level. Brain studies indicate that when we observe someone perform an intentional action, such as swinging a baseball bat or reaching for a cup of coffee, the same neurons in our own brain are activated as if we had performed the action. These "mirror neurons" allow us to interpret the actions of others; we read their minds by (neurally) mimicking their actions and discerning their intentions. This is sometimes called a "perception–action" model of empathy because perception and action—perceiving another's action and performing the action ourselves—rely on similar neural circuits.

Other researchers return to phenomenological characterizations of empathy and argue that *perception* is the primary mechanism for empathy. They

argue that our interactions with others are simply too fast, automatic, and flexible to involve the conscious use of either lay theories or imaginative simulations. There is no phenomenological evidence that we deliberately employ either of these mechanisms to understand others, except in unusual circumstances. Rather, we directly see mental phenomena in ongoing patterns of expressive behavior and respond accordingly. This behavior, as well as an appreciation of the different situations that contextualize it, provides sufficiently rich information to discern others' thoughts, emotions, and intentions. Even if mirror neuron activity is part of this story, it is not the exclusive locus of empathy; rather, it is an aspect of a more complex, temporally extended pattern of social perception. This approach, with its emphasis on face-to-face encounters, is sometimes called a "direct perception," "enactive," or "interactionist" approach to empathy.

Empathy and Affectivity

Another area of disagreement concerns the relation between empathy and affectivity. When we see another's facial expressions and behavior, we often *feel* something in response. A remaining issue concerns the extent to which empathy involves feeling—specifically, feeling *what* another person is going through or minimally feeling a response *to* what he or she is going through. There are two questions here. The first is the extent to which affectivity is necessary for empathy. The second is the question of what compels us to consider and respond with care to the suffering of others. Although the second question deals with important issues concerning motivation and moral psychology, these issues are beyond the scope of this entry.

With respect to the first question, some theorists mark a distinction between empathy and *sympathy*. The former is our basic ability to detect and interpret another's emotions and behavior; the latter is our ability to feel with another person, to either replicate her or his emotion or feel something similar. Empathy—whatever its mechanistic basis—thus proceeds independently of other-oriented affectivity. For example, I can attain knowledge of my friend's suffering via her facial expressions, behavior, utterances, and so on, as well as understand the source of her suffering and how it will guide her behavior, while failing to feel congruent suffering myself. Indeed, becoming highly emotionally aroused can actually impede one's ability to read another's behavior and respond appropriately. So it's not clear that other-directed feeling is necessary for empathy.

Nevertheless, affectivity might be an important part of empathy's ontogenesis. Evidence from developmental psychology suggests that emotional intimacy in early infant–caregiver interactions may provide the developmental context for basic social-cognitive abilities. From birth, the emotional character of these interactions has a motivational effect on infants, compelling them to engage with others. It also guides the infant's attention toward socially salient phenomena such as facial expressions and gestures. *Affect attunement*—the ability to coordinate one's affective states with those of another—thus appears to be the basis from which various individualistic capacities (e.g., self-consciousness, language, self-regulation, self-representation) arise. Since an individualized sense of self and subjectivity is necessary for empathy—that is, to appreciate another's experience *as another's*—affectivity in social relations might be thought of as the ground of empathy, even if affectivity need not be present for empathy to occur.

Judging from the state of current research, the term *empathy* serves as an umbrella term for a relatively heterogeneous group of structures and processes that facilitates different facets of sociality and interpersonal understanding. As such, it is not clear that any single paradigm or disciplinary perspective will adequately capture all of its aspects.

Joel Krueger

See also Embodied Cognition; Folk Psychology; Mirror Neurons and Motor Cognition in Action Explanation; Simulation Theory; Social Cognition; Social Neuroscience; Social Perception; Theory Theory

Further Readings

Decety, J., & Ickes, W. (Eds.). (2009). *The social neuroscience of empathy.* Cambridge: MIT Press.

Gallese, V. (2001). The "shared manifold" hypothesis: From mirror neurons to empathy. *Journal of Consciousness Studies, 8*(5–7), 33–50.

Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading.* Oxford, England: Oxford University Press.

Iacoboni, M. (2008). *Mirroring people: The science of empathy and how we connect with others.* New York, NY: Picador.

Krueger, J. (2012). Seeing mind in action. In Empathy and intersubjectivity [Special issue]. *Phenomenology and the Cognitive Sciences, 11,* 149–173.
Scheler, M. (1954). *The nature of sympathy* (P. Heath, Trans.). London, England: Routledge & Kegan Paul.
Stein, E. (1989). *On the problem of empathy* (W. Stein, Trans.). Washington, DC: ICS.
Stueber, K. R. (2006). *Rediscovering empathy: Agency, folk psychology, and the human sciences.* Cambridge: MIT Press.
Thompson, E. (2001). Empathy and consciousness. *Journal of Consciousness Studies, 8*(5–7), 1–32.
Zahavi, D. (2010). Empathy, embodiment and interpersonal understanding: From Lipps to Schutz. *Inquiry: An Interdisciplinary Journal of Philosophy, 53*(3), 285–306.

Empiricism

The term *empiricism* is often taken to be a vague term for a cluster of doctrines held by the classic modern empiricist philosophers John Locke (1632–1704), George Berkeley (1685–1753), and David Hume (1711–1776), in opposition to the "rationalist" tradition of René Descartes, Baruch Spinoza, and G. W. Leibniz. This takes the terms to be similar to labels for certain tendencies with no clear issue to distinguish them. The result is that many philosophers think these terms are best avoided in serious discussion. But this is premature. There are two clear issues that separate empiricists from rationalists: (1) whether there are innate ideas and (2) whether any propositions can be rationally justified independently of experience. *Conceptual empiricism* is the doctrine that there are no innate ideas but all concepts are acquired through experience, either by introspection or by sense experience, while *justificatory empiricism* holds that all knowledge that something exists independently of experience must be justified on the basis of experience, that is, it must have an empirical justification. The first is a psychological claim about the origin of ideas, while the second is a thesis about evidence and is thus an epistemological claim. The negations of them are conceptual and justificatory rationalism, respectively.

This entry focuses on the British empiricists, in relation to these two issues, innateness and justification. Concept innateness, or *nativism* as it is now called, has come back as one of the central issues debated in contemporary philosophy of mind (this issue is discussed in another entry but indicated here in some items in the Further Readings).

The Two Issues

Concepts

The psychological issue arises in the work of René Descartes (1595–1650). He holds that certain basic concepts such as those of God, infinity, extension, and the mathematical entities cannot be derived from experience and so must be innate. The concept of a line, for instance, cannot come from sense experience, since the lines we experience are only imperfect copies of the lines studied by geometry. Seeing lines on a piece of paper triggers our innate ability to form the concept of the ideal line of geometry, but it does not give us the *idea.* It is an idea of reason or intellect distinct from an image or any idea we can acquire by perception. John Locke, however, denied this and held that the concept derives from experience by abstraction, where experience includes reflection and sense perception, or "inner" and "outer" sense as he calls it. We see lines and leave out their particular characteristics, and by abstraction, we form the notion of a line as an extended series of points with no width. Philosophers discussed this question in the 19th century, but interest in it waned in the early 20th century, when psychology became experimental. The work of the linguist Noam Chomsky reawakened interest in the issue in the later part of the century. Before turning to this, let us look at the question raised by the second issue, that of justification.

Justification

Since Plato, it has seemed clear that knowledge cannot be reduced to true belief. As Socrates argues in Plato's dialogue *Meno,* one might have a true belief about the road to Larissa but be right by accident. He says that beliefs are like the statues of Daedalus, which are so real that they move around. Like them, beliefs need to be "tied down" so they do not "roam about," and the way to do this is to support them with a *logos* or "account." Without this, they cannot qualify as knowledge even if they are true. The modern term for this third element is that the believer must have some reason for thinking that the belief is true, that is, it must be "justified."

The justificatory empiricist holds that if the belief is about a real existent—that is, an entity that exists independently of what is thought about it—it must be justified by sense experience or introspection, while the rationalist holds that claims about real entities can be justified a priori, that is, on grounds independent of experience. Taking *empirical* to mean that the proposition is justified by appeal to experience, we may sum this up by saying that the justificatory empiricist holds that *all knowledge of real existence is empirical,* while the rationalist holds that not all of it is but some is a priori or justified independently of experience.

By "knowledge of real existence" here is meant knowledge that an object or class of objects has real existence, that is, knowledge of truths of the form "There are Xs." Propositions of the form "There are no Xs" may be known empirically (e.g., "There are no two-headed snakes") or a priori (e.g., "There are no round squares"). Experience is not necessary to be justified in believing that there are no round squares, since they are contradictory. The empiricist holds that for us to know that objects have real existence, experience is necessary, but we may know a priori that an object, at least one of a certain special kind (e.g., an impossible object, like a round square), does not exist.

This point is important for understanding George Berkeley. It is sometimes thought that he is not an empiricist since he claims to know that there are no material objects by means of a purely conceptual argument —that is, a priori. Briefly, his argument is that it is inconceivable for something to exist unperceived, since in thinking about it we are perceiving it; that is, he holds that "Material objects exist" is contradictory. For material objects, Berkeley says, "*esse est percipi*" ("to be is to be perceived"). Whether this argument succeeds is a question for metaphysics and need not concern us here (since we are confined to the epistemological side). The important point is that giving an a priori argument for the impossibility of material objects is not inconsistent with his being an empiricist.

More generally, it should be noted that the justificatory empiricist (or just "empiricist" from now on) need not deny that we have a priori knowledge. He might hold that propositions such as that triangles are three-sided figures and 2 + 3 = 5 are a priori but are about useful fictions that have no existence independently of what is thought about them. Locke, Berkeley, and Hume all accepted a priori knowledge, but since they were nominalists (repudiating the reality of universals), this was consistent with their claiming that knowledge of real existence is empirical. Locke, for instance, holds that "The angles of a triangle equal 180 degrees" is a priori (although he never uses the word) and that it is about triangles in nature only if we also know empirically that there are triangular objects. The result is that pure mathematics is a priori but not about real existence, while applied mathematics is about reality but is empirical. Other empiricists, such as J. S. Mill in the 19th century and Willard Quine in the 20th century, go further and reject a priori knowledge altogether, but this does not make Locke any less an empiricist.

Several other points should be noted about this characterization of the issues. First, on this account, empiricism and rationalism on both issues are contradictories. One cannot consistently be both (an empiricist and a rationalist) on either the psychological or the justificatory issue. Second, the justificatory empiricist need not have any opinion on the psychological issue since justification is a matter of evidence and is independent of the genesis of our ideas. Hume appears to be a case in point. He is clearly an empiricist since he holds that all existence claims are empirical, but he also seems to think that the debate over innate ideas is frivolous (*Enquiry Concerning Human Understanding,* sec. II). A third point is that anyone who accepts the ontological argument for God—in other words, the argument that God's existence is known by a priori examining the concept of God—is a justificatory rationalist, while all empiricists must reject it. Finally, these definitions have implications that startle (and perhaps offend) some philosophers. One is that Immanuel Kant is a rationalist since he holds that Newton's law of universal gravitation is a *synthetic* a priori truth about reality. Another is that the early Bertrand Russell is also a rationalist. Russell (1912) says the empiricists were right that all knowledge of existence is empirical, but he also holds that we have a priori knowledge of "subsistent" entities that "have being" timelessly in a Platonic world divorced from ours (*Problems of Philosophy,* chaps. VII and IX).

Criticism and Rejoinders

Empiricism and Innate Ideas

Let us first look at Locke's criticisms of the doctrine of innate ideas and then turn to the issue of

justification. Descartes did not hold that we are always thinking of our innate ideas but that they exist as mental dispositions and need experience only to bring them to consciousness. As he says, children have these ideas in the same way as adult humans have ideas "when they are not attending to them" (letter to Hyperaspistes, 1641). Unlike other mental dispositions (e.g., the ability to speak a language or to play chess), experience is not necessary to first shape them. Let us call them innate *abilities* as opposed to *capacities* (i.e., *dispositions to acquire abilities*). More generally, innate ideas are unstructured powers. An iron bar has the unstructured power to attract iron filings, but it must first be magnetized (or "structured"), perhaps by hitting it with a hammer or passing it through an electrical field. Descartes's view is that our idea of God and the "simple natures" of mathematics are unstructured, innate powers to think of their objects. Experience is only necessary to bring them to consciousness.

One of Descartes's arguments derives from Plato. As we saw, he holds that we cannot acquire the ideas of pure geometry from sense experience since the lines and figures we perceive are not perfectly straight, circular, or triangular. Locke argues that the more plausible theory is that we arrive at them by noting similarities between sensed objects and abstracting from them. They are "perfect" because we ignore the imperfections in our images. Furthermore, the results of this process are fictions or ideal entities (*entia rationis*) that have no real existence outside of thought. Leibniz offered several arguments against Locke in the preface to his *New Essays Concerning Human Understanding,* but most of them rest on misunderstandings of Locke, although this is not generally recognized. First, Leibniz ignores Locke's distinction between inner and outer sense and discounts the role of abstraction. The result is that he merely repeats the traditional arguments that ignore Locke's innovation. Second, Leibniz does not recognize Locke's argument that the innateness doctrine makes all ideas innate if it claims only that we have an innate power to create innate ideas, since every idea we have presupposes an innate power to have it. That is, he ignores what I have called the distinction between capacities and abilities. He bases his criticism on Locke's claim that the mind is a blank tablet at birth (*tabula rasa*), claiming that it is more like a marble block with veins in it. But all Locke meant was that there are no ideas in the mind at birth; he was not denying that we have innate capacities. As Quine was later to put it, the empiricist is "up to his neck in innate mechanisms." The issue is whether abstraction and experience in Locke's expanded sense offer a more plausible explanation of mathematical and logical concepts than innateness, and Leibniz does not address this question.

Historically, conceptual empiricism spurred psychological research in the 19th century through the work of Johann Herbart and Hermann von Helmholtz in Germany and the Mills (John and John Stuart) and Alexander Bain in Britain. Most of this work was speculative or "literary" psychology until Wilhelm Wundt discovered experimentally that the recognition of symbols rests on learning and unconscious processing. In one experiment, subjects were asked to guess three-letter words flashed on a screen too briefly for them to be recognized consciously. He found that native speakers of the language of the target words guessed correctly far beyond what chance would predict, while speakers of other languages performed according to chance. The upshot of this research was that introspection cannot determine when a conceptual ability is innate or derived from experience.

Recently, the linguist Noam Chomsky has revived the debate by arguing that something like the Cartesian doctrine is necessary to explain the fact that humans are only capable of learning a limited range of natural languages. Once again, empiricist critics have argued that this only means that we have an innate capacity to learn only these languages; it does not mean that we have an unstructured innate ability to learn them. It is in this context that Quine remarked that empiricists have always recognized innate capacities. Chomsky has also argued that behaviorism cannot explain language learning since we cannot learn it by generalizing by induction from what we hear. Some behaviorists might overemphasize induction in this connection, but the empiricist is not committed to induction alone. He can argue that children learn language by hearing certain patterns, forming hypotheses, then rejecting them on further experience until they acquire one of the small number of natural grammars the human brain is innately capable of learning. These debates cannot be pursued here, but it seems clear that further progress rests on recognizing the distinction between innate capacities, which need to be shaped by experience,

and innate abilities, which are ready to go without assembly by experience.

Empiricism and Justification

This issue is obscured by the difficulty of getting a clear explanation of a priori justification. Descartes holds that propositions such as "Nothing can be *F* and not be *F* at the same time" and "2 + 3 = 5" are immediately known independently of any sense experience and so must be justified by some other faculty, which he calls "the light of reason" or "pure intellect." But this merely puts a name to the source without explaining it. All we know about it is what it does, not what it is or how it operates. This is similar to saying that opium causes sleep because of its soporific power. By contrast, we have some understanding of experience as a source of evidence from common sense and our knowledge of our sense organs and how they work.

Ironically, when rationalists try to characterize the a priori positively, they always end up comparing it with sense experience. Russell, for instance, says that experience may make us think of the proposition that 2 + 2 = 4 but "it does not suffice to prove it"; it merely directs our attention so that "we see its truth without requiring any proof from experience" (*Problems of Philosophy,* chap. VII). "Seeing" here means seeing with the mind and is a metaphor. Our conviction about the proposition resembles what we have when we see with our eyes that there is an apple on the table, but it is intellectual perception. The problem is that nonsensory seeing does not provide a mark to distinguish it from intense conviction that is not in itself evidential. It may appear to Russell to be an act of insight on par with seeing an apple, but it may also be just a firm belief. It is surely not uncommon to confuse conviction with insight. It is sometimes said that we "grasp" the truth of propositions like "2 + 2 = 4" as soon as we understand them, but this is another metaphor. These metaphors can be traced to the ancient Stoic doctrine of the *kateleptic* (from the Greek word *kataleipsis,* meaning "taking hold of," "grasping," or "seizing"), that is, the grasping cognition that certainty is like a clenched fist that holds its object tightly so it cannot escape, but conviction is also a clenched fist. Sometimes the issue reduces to name-calling. In the phenomenological tradition, Edmund Husserl (1907/1999) says, for instance, that a priori cognition is directed at "general essences" and is a "luminous intuition." The skeptic who denies such cognition is like a blind man who denies that there is such a thing as seeing. "How could we convince him, assuming that he has no other mode of perception?" (*The Idea of Phenomenology,* Lecture IV). But this is not helpful; the skeptic will reply that arguing with Husserl is like talking to a madman who is prone to hallucinations. The empiricist's conclusion from this is that a priori justification is hopelessly obscure.

One curious feature of this debate is that the concept of an innate idea seems clearer than that of the a priori itself. As we saw, Descartes thinks that the idea of a triangle is in the mind from birth in the same way in which the ideas of familiar objects are in the mind when we are not actually thinking of them. We can bring an innate idea to consciousness by attention alone and form judgments about it without prior experience. He may be mistaken that we have such ideas, but at least it is clear what he means by them.

The obscurity of the a priori also affects moderate empiricists, like Locke, Berkeley, and Hume, who accept a priori knowledge but deny that it informs us about real existence. But there is also a difference. Moderate empiricists are willing to accept the a priori since it only covers propositions that are true according to the meanings of their terms (e.g., analytic sentences), and these are based on convention alone, while substantive claims about reality can be known only on the basis of experience. The rationalist, on the other hand, wants the a priori to give knowledge about real existence. Sometimes he claims that essences do not exist but subsist, but whatever he calls them, they are not *entia rationis* but have an independent existence, and this is what the empiricist finds suspect.

The main argument for the a priori is that it is needed in order to answer the skeptic about the general principles of knowledge. This can be illustrated by the problem of induction. Hume (1748/2007) holds that we know generalizations such as that bread nourishes because it has nourished us in the past and we believe that "like causes will be followed by like effects" (*An Enquiry Concerning Human Understanding*). But this principle is itself a matter of fact, and his empiricism commits him to holding that all knowledge of matters of fact must be justified by experience. Hence, any attempt to justify it by experience would be circular. This is known as the problem of induction. Hume thinks that the problem is unsolvable but that we cannot

help accepting induction; we believe the principle by instinct, or what he calls custom, and "nature is too strong for principle." The rationalist argues that the problem shows empiricism to imply that we have no knowledge by induction. For an inferential practice to yield justified beliefs, we must be able to justify the principles on which it operates, and the only way to do this in this case is by appealing to the a priori.

This debate is ongoing, but two points might be ventured here in closing. First, it is not clear how appealing to the a priori will solve the problem, given its obscurity. Second, and more interestingly, it is not clear that we must justify induction in order to have inductive knowledge. All reasoning must rest on premises that are either accepted as ultimate or justified immediately by something like Husserl's "luminous intuition," which, as we saw, cannot be clearly distinguished from strong conviction. The result is that perhaps the most we can hope is that Hume's principle is true. Not being able to justify it noncircularly is just one facet of the human condition. To show that not being able to justify it noncircularly leads to skepticism, the rationalist must give a convincing argument that it does, and this is not promising given the obscurity of the a priori.

Robert G. Meyers

See also A Priori and A Posteriori; Analytic/Synthetic Distinction; Concepts; Epistemology; Experimental Philosophy; Idealism; Induction and Confirmation

Further Readings

Bennett, J. (2003). *Learning from six philosophers: Descartes, Spinoza, Leibniz, Locke, Berkeley, Hume* (Vols. 1 & 2). Oxford, England: Clarendon.

Berkeley, G. (2008). *A treatise concerning the principles of knowledge*. In D. M. Clarke (Ed.), *Berkeley philosophical writings* (pp. 67–150). Cambridge, England: Cambridge University Press. (Original work published 1710)

Boring, E. G. (1957). *A history of experimental psychology*. New York, NY: Appleton-Century-Crofts.

Carruthers, P., Laurence, S., & Stich, S. (Eds.). (2007). *The innate mind*. Oxford, England: Oxford University Press.

Chomsky, N. (2006). *Language and mind* (3rd ed.). Cambridge, England: Cambridge University Press.

Cowie, F. (1999). *What's within: Nativism reconsidered*. Oxford, England: Oxford University Press.

Descartes, R. (1985). *The philosophical writings of Descartes* (Vols. 1–3). Cambridge, England: Cambridge University Press.

Hume, D. (2007). *An enquiry concerning human understanding*. Oxford, England: Oxford University Press. (Original work published 1748)

Husserl, E. (1999). *The idea of phenomenology*. Dordrecht, Netherlands: Kluwer Academic. (Original work published 1907)

Kaldis, B. (2011). Leibniz's argument for innate ideas. In M. Bruce & S. Barbone (Eds.), *Just the arguments: 100 of the most important arguments in Western philosophy* (pp. 283–289). Oxford, England: Wiley-Blackwell.

Leibniz, G. W. (1996). *New essays concerning human understanding* (P. Remnant & J. Bennett, Eds.). Cambridge, England: Cambridge University Press. (Original work published 1765)

Locke, J. (1979). *An essay concerning human understanding* (P. Nidditch, Ed.). Oxford, England: Clarendon. (Original work published 1690)

Meyers, R. G. (2006). *Understanding empiricism*. Chesham, England: Acumen.

Priest, S. (2007). *The British empiricists* (2nd ed.). London, England: Routledge. (First published 1990)

Russell, B. (1912). *Problems of philosophy*. London, England: Oxford University Press.

Stich, S. (Ed.). (1975). *Innate ideas*. Berkeley: University of California Press.

Woolhouse, R. S. (1988). *The empiricists*. Oxford, England: Oxford University Press.

Encyclopedia

Encyclopedia has been an important and enduring form of the human endeavor toward an exhaustive, universal exposition of knowledge, to which both scientific knowledge and philosophical knowledge have aspired.

This entry introduces the origin of the word *encyclopedia* and presents the formal elements that make up the genus "encyclopedia" as well as its more specific typology. The entry discusses these features, first generally and then more specifically, in their philosophical dimension. The entry also includes a brief reference to the more significant historical moments of *encyclopedism* as well as to the actual developments of encyclopedia, namely, electronic and online encyclopedias, hypertext, and the Internet as its actual realization.

The Word

The word "encyclopedia" (Greek ἐγκυκλοπαίδεια) comes from the Hellenistic expression *enkuklios paideia* (ἐγκύκλιος παιδεία) meaning "perfect or regular cycle of studies"—in other words, an instituted general course (or cycle) of studies. It is taken up again in Roman culture, appearing in the dedication to Emperor Titus in the *Historia Naturalis* of Plinius. In early-modern times, the word was first used in English by Sir Thomas Elyot (*Boke of the Governour,* 1531) and in French by Rabelais (*Pantagruel,* 1532).

There are many works that may be retroactively considered as belonging to the genus encyclopedia, although they had not carried that designation in their title (e.g., Vincent de Beauvais's *Speculum Majus,* ca. 1260). In fact, it is only in the 17th century that the word *encyclopedia* comes close to its current meaning. It is also about then that as a literary product encyclopedia becomes a kind of its own, bearing its distinctive identity and characteristic procedures.

The Elements of an Encyclopedia

Encyclopedia aims, ideally, to be the exposition of *all* knowledge conquered by mankind. Sometimes such aims toward an exhaustive inclusion of everything known may lead encyclopedias to a teratological dimension, as in the case of the immense Chinese encyclopedias. However, in the West, encyclopedias are of a rather contained or relatively limited kind, combining comprehensiveness with selectivity.

Encyclopedia is also a *historical production* reflecting specific cultural and scientific conditions. Never complete, always destined to be surpassed by ongoing progress in knowledge, it requires constant reworking or renewal. Such an updating consists of not only including new entries according to recent developments (e.g., "Cloning") but also diminishing the weight of some entries or altogether deleting others (e.g., "Phlogiston") and raising the importance of yet other entries (e.g., "Globalization").

Encyclopedia is *not a dictionary.* Dictionaries aspire to be a full, consistent codification of language—that is, they are contained within linguistic meanings. On the contrary, encyclopedia is semantically open, beyond mere meanings, to the actual things, events, theories, or concepts to which the terms it contains refer, that is, to what is known. Although many encyclopedias have been designed as dictionaries, an encyclopedia is never concerned with words but with what specific terms or concepts mean or refer to: the world behind the words.

Like dictionaries, encyclopedias are a discontinuous text made of independent entries, which, however, never present well-defined borders. Each entry opens (explicitly or implicitly) to other entries (cross-references or "See also"), which in turn open to others, in such a way that each entry is virtually connected with all the others. Thus, encyclopedia is not so much a mereological aggregate of all items of knowledge placed, in a disconnected fashion, in one enclosed *topos* but rather a circulation (*navigation*) throughout the vast net of its multidimensional elements.

Thus, the material objectivity of encyclopedia has an *unlimited and combinatory condition*. Behind the additive synthesis of all its entries, encyclopedia points to the exhaustion of all the possible combinations of its entries, which can be articulated in an undetermined number of combinations—a kind of combinatory without rule.

This situation has important consequences for the *reading regime* of encyclopedia. The readers of an encyclopedia are offered the possibility of making their own journey of reading according to their interests and preferences (surfing). It is the reader (navigator) who selects, by successive choice, which semantic field she or he will read further on. In fact, encyclopedia affords this possibility, proposing several resources that invite the reader to actualize one of the many possibilities it offers (indexes, thesaurus, internal references, reader's guide, cross-connections, articulations, and instantaneous electronic *links,* which today characterize hypertext and the Internet as its extension).

Encyclopedias extensively use *nonlinear materials* such as maps, drawings, illustrations, statistic schemata, plans, and tables of all types (see the 11 complementary volumes of tables published by Diderot or the wide number of imaging resources in electronic encyclopedias, in hypertexts, and on the Internet). That is to say, all encyclopedias point to the semantic exploration of the diagrammatic resources of language, putting them in the service of the iconic, imagistic, cartographic description of the world.

Encyclopedia is a *collective work*. Some encyclopedias were made by only one author (e.g., Pierre Bayle's *Dictionnaire Linguistique et Critique,* 1647–1706). However, encyclopedia inevitably presupposes the collaboration of different competencies:

celebrated scientists, appointed experts, identified authors, as well as anonymous, unidentified, and even unknown collaborators (as in today's *Wikipedia*).

Implicitly or explicitly, encyclopedia is always an *ordered, organized structure*. It is that ordered structure that determines the quantity and quality of entries, the inclusion or exclusion of certain topics, and the relative importance of some entries compared with others. This does not mean that encyclopedias should be endowed with a unique, constraining point of view. What it means is that encyclopedias have an undeniable synthesizing and structuring task.

Encyclopedia always has a *prospective role*. By promoting terminological unification, local systematizations, and interdisciplinary connections, encyclopedia favors the establishment of bridges from one domain to another. By synthesizing what is already known, encyclopedia constitutes a kind of mnemonic prosthesis, which liberates or unburdens natural memory for what really matters: the unknown. It intends to reduce the opposition between memory and invention. That is why encyclopedia is such a strong anti-Cartesian device. It does not make a *tabula rasa* of previous achievements.

Typology of Encyclopedia

Encyclopedias are essentially *general in kind*, offering an (as much as possible) exhaustive yet concise and selective exposition of the empirical, scientific, or philosophical knowledge acquired by humankind until a given point in time. They may be *disciplinary* in kind (based on a classification of sciences), *alphabetic* (operating by definition of a huge set of concepts alphabetically ordered), *thematic* (constituted by large entries covering a set of selected subjects or whole fields), or *mixed* (aiming to safeguard the combined benefits of disciplinary and thematic encyclopedias with the practical advantages of the alphabetic ordering).

However, confronted with the development of scientific knowledge and its compartmentalization as well as with a growing cultural complexity, encyclopedias were forced, mostly during the 19th century, to become *specialized*, focusing on a single or specific human activity or domain of knowledge.

Philosophical encyclopedias are such a type of specialized encyclopedias, aiming to cover a certain philosophical tradition or branch. Nevertheless, they may be considered to coincide with the very philosophical project as such. The case of Gottfried Leibniz in the 17th and 18th centuries and of Otto Neurath in our own times will serve to illustrate this. For both philosophers, it is through encyclopedia that philosophy reaches its aim of a *unified* knowledge of the world.

It is true that Leibniz did not provide any actual encyclopedia. He did, however, formulate with extreme rigor the theoretical presuppositions and problems associated with such an enterprise. He clearly understood the need and heuristic significance of constructing, in parallel, both the encyclopedia and the *characteristica universalis*. If the construction of the scientific language presupposes the analysis of all fundamental concepts and their definitions—that is, the encyclopedia—then, on the other hand, the symbolic transposition of the cognitive conceptual contents of the linguistic system that is already constituted allows for the discovery of new possible meanings, thus favoring the accomplishment of encyclopedia. By revealing the diverse regions of science, encyclopedia facilitates the analytical decomposition of terms. But, inversely, the use of a symbolic system capable of expressing the various ideas and their relationships leads to the full analysis of those ideas and, thus, to their definition and systematic articulation; that is, it helps the construction of encyclopedia.

As for Neurath, he refused the need of both a foundationalist basis for encyclopedia and its unique, total, a priori systematic organization. He emphasized the cooperative, historical, and unfinished nature of encyclopedia, which he considered to be the symbol of the scientific fraternity and a living intellectual force in the service of mankind. Like Leibniz, Neurath gave great emphasis to the heuristic role of encyclopedia, which he considered to be a kind of *organon* for scientific progress. In doing so, he fully identified encyclopedia as the "model" of human knowledge and as the main task for the *unity* of science. Such unity, like encyclopedia itself, should be plural, heterogeneous, always provisory, and incomplete while open to further cooperative contributions.

Historical Summary

From the *Historia Naturalis* of Plinius (Caius Plinius Secundus or Plinius the Elder, 23–79 CE) to the *Etimologies* of Isidorus (St. Isidore of Sevilla, ca.

560–636) or the *Eruditionis Didascalicae* of Hugh (Hugh of Saint Victor, ca. 1096–1141), it is possible to talk of a medieval and even Roman encyclopedia. However, in a more precise sense, encyclopedia is a product of the 18th century. Facing the intense scientific activity of modern times, encyclopedia appears as the *topos* where new forms of disciplinary and systematic arrangement of knowledge and new schemes of intelligibility are intended (e.g., Bacon's *Instauratio Magna,* 1620; Alsted's *Encyclopaedia Omnium Scientiarum,* 1630; Comenius's *Pansofia Prodomus,* 1639; as well as Leibniz's "Mathesis Universalis"—dispersed in many fragments and texts of Leibniz's).

The 18th century is the golden age of *encyclopedism*. Projects turned into successful, actualized outputs. Having reached its classical form, science offered itself the luxury of its own monumentality. Encyclopedias started playing the role of a medieval cathedral, a monument in which human beings face what is more important: not God anymore but the created World. The major realization is undoubtedly the *Encyclopédie ou Diccionaire Raisonné des Sciences, des Arts et des Métiers* (1751–1765), by Denis Diderot and Jean D'Alembert, a kind of laboratory where it is possible to grasp not only all the positive qualities but also the difficulties of the encyclopedic project. Diderot and D'Alembert were able to assemble many of the most eminent intellectuals of the century as well as humble artisans of all areas of human activity. They were able to construct a kind of a cartographic overview not of the secrets of the world (of which God was until then the first and final cause) but of the heterogeneous, contingent, and precarious achievements of human knowledge. The Enlightenment encyclopedia of Diderot and D'Alembert is in fact a singular moment in the history of Western culture in which, for the first time, a secular world was envisaged, that is, a reality with its own autonomy and intelligibility was opened to the human intelligence and industry as the background to a further, infinite cognitive conquest.

In the 19th century, two divergent lines of encyclopedism were put forward. Under the positivist regime proposed by Auguste Comte, encyclopedia became the instrument for the coordination of particular sciences. It had to consider the achievements of each discipline and point to the logical law of their development (e.g., *Grande Encyclopédie: inventaire raisonné des sciences, des lettres, et des arts,* by André Berthelot and Ferdinand-Camille Dreyfus, 1886–1903). On the other side, German romantic encyclopedism aimed to construct a "total science" on the basis of the deep analogies among the particular sciences (see Novalis's *Fragmente,* 1802) or to expose their organic unity while being necessary parts of the philosophical whole (as in Hegel's *Enzyklopädie der Philosophischen Wissenschaften im Grundrisse,* 1817).

In the first decades of the 20th century, and in the context of Logical Positivism, Otto Neurath, whom we encountered above, had the merit of having conceived the original plan and of having effectively taken charge of the organization of the *International Encyclopedia of Unified Science,* of which the part called *Foundations of the Unity of Science* was published in 1938 to 1969 in two volumes (comprising in total 19 monographs, among which was Thomas Kuhn's celebrated *The Structure of Scientific Revolutions,* which appeared in 1962 as the second monograph of Volume II).

The decades between 1940 and 1960 were characterized by a generalized disillusion with the very idea of encyclopedia. Above all, in literature, scholars like Alberto Savinio, Georges Perec, Italo Calvino, or Jorge Luis Borges came to explore the paradoxes of any such totalizing attempt and to announce the end of any encyclopedic effort.

However, a surprising renewal was to take place in the 1970s. Contrary to what might have been expected, encyclopedia took up a new integrative format intended to critically prevail over the huge fragmentation of scientific knowledge and information. The most innovative encyclopedias—namely the *Encyclopedia Universalis* (1968–1975) and the *Encyclopedia Einaudi* (1977–1984)—put into practice thematic, interdisciplinary structures dealing with controversial matters, pioneering concepts, current debates, and unexpected articulations.

Current Developments

In the last decades of the 20th century, another decisive development took place. Hypertext, devised by Ted Nelson (1965), and the World Wide Web, developed by Tim Berners-Lee (1991), can directly be inscribed in the history of encyclopedia. That is why, as soon as the necessary technical conditions were in place, the classical form of encyclopedia promptly adhered to its own informatization (*Britannica's* first

CD-ROM and *online* editions appeared in 1993 and 1994, respectively). This development can be seen as if the electronic encyclopedia and, further, the hypertext and the Internet, all came to realize the primordial aspirations of encyclopedia itself.

Of course, in the case of hypertext—which can be taken as an actual apex of the encyclopedia proper—there are enormous difficulties, terrible noise, a dangerous lack of cartographic references, serious deficiency of selectivity and credibility of contents, and so forth. But there are enormous advantages as well. Like encyclopedia, the hypertext is a generalist technology of knowledge operating by a regime of multiple choices among always growing yet never fully actualized, virtual contents. And the access to all those possibilities is effectuated by the simple activation of a mouse's click.

Occupying the privileged place of the unity of science, encyclopedia and today's hypertext are a powerful form of resistance against the effects of specialization. Encyclopedia and hypertext are the only attempts toward unification of knowledge that can claim success in being effectively accomplished. They are the only forms of a material realization of the goal of the unity of science that condense and present to the eyes of all of us a large part of that information that could never be seen, grasped, and confronted otherwise.

Olga Pombo

See also Collective Memory; Embodied Cognition; Enlightenment, Critique of; Information Society; Interdisciplinarity; Knowledge Society; Neurath's Unity of Science and the Encyclopedia Project; Philosophes, The

Further Readings

Blair, A. (2010). *Too much to know: Managing scholarly information before the modern age*. New Haven, CT: Yale University Press.

Cherchi, P. (1990). *Encyclopedism: From Pliny to Borges*. Chicago, IL: University of Chicago Press.

Collison, R. (1966). *Encyclopaedias: Their history throughout the ages*. New York, NY: Hafner. (A bibliographical guide, with extensive historical notes, to the general encyclopedias issued throughout the world from 350 BCE to the present day)

Giorello, G. (1973). Le Systéme des Savoirs [System of knowledge]. In *Encyclopaedia Universalis: Supplement II, Les Enjeux* [The issues] (pp. 553–559). Paris: Encyclopaedia Universalis France.

Pombo, O. (2011). Neurath and the encyclopaedic project of unity of science. In O. Pombo, J. Symons, & J. M. Torres (Eds.), *Otto Neurath and the unity of science* (pp. 59–70). Dordrecht, Netherlands: Springer.

Pons, A. (Ed.). (1963). *Encyclopédie ou Dictionnaire Raisonné des Sciences, des Arts et des Métiers* [Encyclopedia or a systematic dictionary of the sciences, arts and crafts]. Paris, France: Bussière.

Salzano, A. (1973). Encyclopédie et Encyclopédisme aujourd'hui [Encyclopedia and encyclopedism today]. In *Encyclopaedia Universalis: Supplement II, Les Enjeux* [The issues] (pp. 560–568). Paris, France: Encyclopaedia Universalis France.

Salzano, A. (1977). Enciclopedia [Encyclopedia]. In *Enciclopedia Einaudi* (Vol. 1, pp. 3–62). Turin, Italy: Guilio Einaudi Editore.

Enlightenment, Critique of

This entry presents the critical voices the Enlightenment attracted against itself both during its heyday and more recently, when what is known as "the Enlightenment project" became the subject of critical controversy. The tenets of Enlightenment have been vital both to political philosophy and to the rise of modern social sciences. Thus, a critical turning against the Enlightenment signals a critical turning, too, against received views about the nature of social science.

Origins

The Enlightenment and its critics were born together, like twins, in the middle of the 18th century. It was then that a group of thinkers in Europe and the United States first began to view themselves as part of a very broad, loose movement of reform committed to reason, science, and religious toleration. In France, where this movement was most self-conscious, they were referred to, and referred to themselves as, *philosophes,* a word for which there is no precise English equivalent. In French, it means both philosophers in the narrow sense and, more broadly, public intellectuals and thinkers, including natural scientists, clerics, politicians, public officials, novelists, and journalists. They also sometimes referred to themselves collectively as a "society of

men of letters" who owed allegiance to a cosmopolitan "republic of letters" devoted to reform and the popular dissemination of enlightenment. While the best-known *philosophes* were French—men such as Voltaire, Denis Diderot, Jean d'Alembert, the Comte de Buffon, Étienne de Condillac, and Baron de Montesquieu—they had counterparts across Europe and the United States, most notably Adam Smith and David Hume in Scotland, Immanuel Kant and Gotthold Lessing in Germany, Thomas Jefferson and Benjamin Franklin in the United States, and Cesare Beccaria in Italy. Virtually all the *philosophes* in France contributed to the *Encyclopédie,* a vast compendium of knowledge that comprised 28 large volumes with more than 70,000 articles and illustrations that served as a *machine de guerre* of the Enlightenment designed to "change the way people think," in the words of its editor, Diderot.

Early Phase

Perhaps the first significant opponent of this movement was one of its own members, the Swiss writer Jean-Jacques Rousseau, a defector from within its own ranks who had been a friend and ally of the *philosophes*. Yet in his *Discourse of the Sciences and the Arts* (1750), he praises ignorance and associates the acquisition of knowledge of the arts and sciences (modern "civilization") with decadence and moral depravity. Many *philosophes* were shocked by this argument, including the coeditor of the *Encyclopédie,* d'Alembert, who rebuked Rousseau in his *Preliminary Discourse to the Encyclopédie* (1751), which became the movement's unofficial manifesto in France. This early skirmish soon escalated into a major clash between Rousseau and the *philosophes* and launched a war between the Enlightenment and its opponents that has raged ever since.

In Germany, Johann Georg Hamann launched a crusade against the German Enlightenment—(*die Aufklärung*)—and its institutional embodiment there, the Prussian state of Frederick II, based in Berlin. Beginning in 1758, he wrote a series of essays attacking the enlightened despotism of Frederick in the name of his own mystical conception of faith, denouncing the pedantic "lettered men of our enlightened century" as dogmatists of a new secular religion whose "bible," the *Encyclopédie,* is a barren substitute for the Holy Bible. His enlightened friend Kant was dismayed and puzzled by Hamann's turn against the *Aufklärung,* just as d'Alembert felt betrayed by Rousseau's attack on the values of the Enlightenment a few years earlier.

In 18th-century France, where the church was formally allied to the state and retained considerable power and influence, many of the earliest opponents of the *philosophes* were conservative Catholic *dévots,* such as Nicolas-Sylvestre Bergier and Guillaume François Berthier, who attacked the *encyclopédistes* in periodicals like the *Journal de Trévoux* for undermining Christianity and for the alleged corrosion of popular morals. Many of these early religious critics of the *philosophes* portrayed them as dangerous atheists blind to the potentially nihilistic consequences of disbelief. This was a convenient caricature, since very few *philosophes* actually denied the existence of God. While some, such as Claude Adrien Helvétius and the Baron d'Holbach, were indeed atheists (and philosophical materialists), they were rare exceptions; most *philosophes* were moderate deists, like Voltaire, who were as critical of atheism as they were of religious fanaticism. When the Faculty of Theology at the University of Paris unanimously passed the heterodox thesis of the Abbé de Prades in 1751, it sparked enormous controversy and opposition (the so-called Prades affair), setting religious traditionalists against anticlerical reformers in France during the decades that followed. The *Encyclopédie* was eventually suppressed in France in 1759, although it continued to be published unofficially with the support of some very high-placed officials, including Guillaume-Chrétien Malesherbes, the director of censorship and publication.

It was not until the French Revolution that criticism of the Enlightenment really took off. Many blamed the violent excesses of the Revolution on the Enlightenment, which it was widely believed had systematically destroyed the legitimacy of the ancien régime and plunged Europe into violent chaos for a generation. Among the most eloquent proponents of this view was the ultramontane Savoyard Joseph de Maistre (subsequently an early icon of conservatism in political thought), whose *Considerations on France* (1796) depicts the events of the 1790s as divine punishment for the sins of the Enlightenment. Its most popular proponent was the Abbé Augustin Barruel, whose bestselling *Memoirs Illustrating the History of Jacobinism* was published in 1797. In it, he makes the case that a deliberate conspiracy

of *philosophes,* Freemasons, and the international Order of the Illuminati plotted to overthrow the ancien régime in Europe with a violent revolution. Edmund Burke, who read and admired Barruel's *Memoirs* shortly before his death, had already traced the origins of the Revolution to the *philosophes* in his enormously successful and influential *Reflections on the Revolution in France* (1790), which indicts Condorcet, Voltaire, Rousseau, d'Alembert, Diderot, and Helvétius by name for destroying the foundations of social and political order in France with their corrosive skepticism.

Many early romantic writers at the end of the 18th and early 19th centuries in France, Germany, and Britain shared in the criticism that the Enlightenment was both politically dangerous and spiritually empty. Many of these writers were very critical of the Enlightenment's allegedly narrow emphasis on reason at the expense of emotion and the passions, which, they claimed, had made the 18th century an age devoid of beauty, imagination, and spirit. This is the central theme of François-René Chateaubriand's *The Genius of Christianity* (1802), a popular aesthetic defense of Christianity, which gives the beauty and mystery of faith a centrality absent in the rationalistic natural religion that was popular among the *encyclopédistes*. He condemns the "atheistic sect" of the *philosophes,* led by Voltaire, for promoting a disenchanted view of the universe that is sterile, ugly, and reductionistic. Similar views were expressed by the German nobleman Georg Friedrich Philip von Hardenberg (who used the nom de plume Novalis). In *Christendom or Europe* (1826), he paints a very unflattering picture of the Enlightenment as a sterile secular age made up of "rigid seas, dead cliffs, fog instead of starry heavens," in stark contrast to the beauty and enchantment of medieval Catholicism, which he depicts as a time of simple beauty and deep piety.

Later Phase

Not all of the Enlightenment's critics were religious conservatives or traditionalists. The Enlightenment became a key organizing concept in much of social and political thought after World War II, in the process attracting large numbers of critics from all points of the ideological compass, including some liberals, socialists, feminists, environmentalists, and postmodernists. For example, the Enlightenment played a central role in the understanding of 20th-century totalitarianism for many intellectuals, such as the Neo-Marxists Max Horkheimer and Theodor Adorno, whose influential 1947 work *Dialectic of Enlightenment* depicts Western enlightenment in all its forms as linked to a narrowly instrumental conception of reason that is inimical to decent human life. Cold War liberals such as Jacob Talmon and Isaiah Berlin also identified the legacy of the Enlightenment as one that was perversely connected to destructive political pathologies in the 20th century, particularly communism. Some "poststructuralists" have echoed these ideas. Michel Foucault, whose views were influenced by Horkheimer and Adorno, wrote several books in the 1960s and 1970s chronicling the emergence in the late 17th and 18th centuries of a deeply sinister "disciplinary society" that was masked and justified by a deceptively liberal and humane rhetoric. The feminist writer Jane Flax has attacked the Enlightenment for its supposedly "pure" conception of reason, which is a form of "androcentrism" that suppresses key gender differences in the interests of men.

One of the most common and consistent themes among the Enlightenment's critics since the 18th century is what can be called its *perversion of reason*—distorted conceptions of reason that many associate with the Enlightenment. It plays an important (in many cases central) part in almost every depiction of the Enlightenment by its critics, from the mid 18th century to the present and from the far left to the far right. Some, like Rousseau and de Maistre, have accused the *philosophes* of greatly exaggerating the power and influence of reason and underestimating the importance of nonrational influences on human behavior, such as conscience, in Rousseau's case, or sin, in de Maistre's. Others, such as Horkheimer and Adorno, have alleged that the Enlightenment had a narrowly instrumental conception of reason, incapable of providing insight into objective truths about morality, justice, beauty, or the proper ends of life and that this conception has turned the West into an "iron cage." Finally, writers such as Hamann and Flax have accused the *philosophes* of propounding a view of reason as "pure," when in fact it is inextricably mixed with power and interests. In each of these cases, the proper role, character, and importance of reason have allegedly been grossly distorted by the Enlightenment, usually with disastrous consequences.

It is important not to misrepresent this conflict as one between friends and enemies of reason, even though many of the Enlightenment's critics have tried to do so. Although objections have consistently been raised against what has been taken to be the "typical" Enlightenment view of reason by its critics, this has almost never been generalized to reason as such. None of the critics of the Enlightenment have been prepared to abandon reason entirely. And most 18th-century *philosophes* were keenly aware of both the power of the passions and the limitations of reason. Very few agreed with Descartes that reason is, or ever could be, absolutely sovereign, and some (e.g., Diderot) defended the passions from the proponents of both Christian dogma and Cartesian rationalism. The battle between the Enlightenment and its critics has really been over the scope, meaning, and application of reason, not over whether it is good or bad, desirable or undesirable, essential or inessential per se.

The history of Enlightenment criticism is as much a history of constructions of "the Enlightenment" as it is of criticisms of it. Although the philosophers of the Enlightenment used the terms *éclaircissement* in French and *Aufklärung* in German to refer to the general process of replacing ignorance with knowledge or insight, just as *enlightenment* was used in English, they did not have a term comparable with "the Enlightenment" (definite article, capital *E*) to refer to a particular historical period. The latter was a retrospective invention, just as "the Renaissance" and "the Reformation" were. "*Die Aufklärung*" was used from the late 18th and early 19th centuries in German to refer to the period we now call "the Enlightenment," not the least by Hegel, whereas "the Enlightenment" did not appear in English until the late 19th century and was not widely used until after World War II. The French term *le siècle des Lumières* (the century of lights) is even more recent. This is significant because it suggests that many later writers have projected back onto the 18th century a unity and self-consciousness that it lacked at the time, often to set up a straw man that could easily be knocked down, attributing a degree of coherence and purpose to the *philosophes* that never really existed. As our view of the internal complexity and diversity of the Enlightenment has become increasingly sophisticated in the past 40 years, many scholars have become wary of referring too casually to "the Enlightenment," let alone to a single "Enlightenment project." Some have even argued that the term should be abandoned as a straightforward anachronism. We have already seen, for example, that very few of the Enlightenment's 18th-century proponents were atheists, even though many of their critics have accused them of being just that and blamed them for the supposed consequences of disbelief.

Another common belief attributed to the Enlightenment is that its advocates were naive optimists who believed in the inevitability of progress, even though Voltaire, in many ways the quintessential Enlightenment figure, openly mocked this view in his popular novel *Candide* (1759). At best, they were very cautiously optimistic about the prospects for improvement, with a keen sense of how slow and uncertain it could be. Even so, most believed that things had improved and would likely continue to do so as reason, toleration, and science gradually displaced religion, intolerance, and superstition.

Not all of the Enlightenment's critics wished to undermine it, let alone destroy it. There are friendly critics sympathetic to its basic values and goals who have rejected some of its assumptions and prescriptions in order to save it, just as some 20th-century Marxists have revised some of Marx's original claims in order to strengthen the general theory in light of our knowledge and experience of the world since the 19th century. For example, Berlin thought that the German philosopher Johann Georg Herder was an early example of this, describing him as not an enemy but a critic of the French Enlightenment. Recently, the philosopher Jürgen Habermas has criticized some aspects of the Enlightenment in order to improve it rather than to undermine it. Such friendly critics should be clearly distinguished from more extreme critics or "enemies" of the Enlightenment, like Hamann and de Maistre.

Graeme Garrard

See also Empiricism; Foucault's Thought; Herder's Philosophy of History; Idealism; Modernity; Montesquieu and the Rise of Social Science; Philosophes, The; Postmodernism

Further Readings

Berlin, I. (2000). *Three critics of the Enlightenment: Vico, Hamann, Herder.* Princeton, NJ: Princeton University Press.

Berlin, I. (2001). The counter-Enlightenment. In *Against the current: Essays in the history of ideas* (pp. 1–24). Princeton, NJ: Princeton University Press.

Garrard, G. (2006). *Counter-Enlightenments: From the eighteenth century to the present.* London, England: Routledge.

Habermas, J. (1990). *The philosophical discourse of modernity.* Cambridge: MIT Press.

Habermas, J. (1997). Modernity: An unfinished project. In M. P. d'Entrèves & S. Benhabib (Eds.), *Habermas and the unfinished project of modernity: Critical essays* (pp. 38–58). Cambridge: MIT Press.

Kors, A. (1976). *D'Holbach's coterie: An enlightenment in Paris.* Princeton, NJ: Princeton University Press.

Lestition, S. (2007). Countering, transposing, or negating the Enlightenment? A response to Norton. *Journal of the History of Ideas, 68,* 659–681.

Mali, J., & Wokler, R. (Eds.). (2003). *Isaiah Berlin's counter-Enlightenment.* Philadelphia, PA: American Philosophical Society.

McMahon, D. (2001). *Enemies of the Enlightenment: The French counter-Enlightenment and the making of modernity.* New York, NY: Oxford University Press.

Norton, R. E. (2007). The myth of the counter-Enlightenment. *Journal of the History of Ideas, 68,* 635–658.

Schmidt, J. (Ed.). (1996). *What is Enlightenment? Eighteenth-century answers and twentieth-century questions.* Berkeley: University of California Press.

Schmidt, J. (2000). What Enlightenment project? *Political Theory, 28*(6), 734–757.

Schmidt, J. (2003). Inventing the Enlightenment: Anti-Jacobins, British Hegelians and the Oxford English Dictionary. *Journal of the History of Ideas, 64*(3), 421–443.

Sternhell, Z. (2010). *The anti-Enlightenment tradition* (D. Maisel, Trans.). New Haven, CT: Yale University Press.

Epistemic Approaches to Democracy

This entry introduces the idea of establishing the value of democracy on the basis of epistemic merits expected to accrue from certain democratic procedures. The discussion of such *epistemic approaches* to democracy blends epistemological issues (including those of rational choice and probability) with political philosophy and political science. The entry critically reviews various approaches and ends by discussing what is known as *Condorcet's jury theorem.*

Epistemic approaches to democracy argue that its value consists at least partly in the tendency of some democratic arrangements to make good political decisions. The name, based on the Greek word for knowledge or wisdom, *epistêmê,* was coined only recently by Jules Coleman and John Ferejohn, though the approach appears throughout the tradition—as far back as in the work of Jean-Jacques Rousseau or more recently in the work of Alexander Meiklejohn. Epistemic approaches are not committed to moral consequentialism, since they might be combined with noninstrumental moral principles of justice or justification.

Democracy has long been burdened by the charge that it is absurd to decide such important matters by neglecting the differences between those who have relevant knowledge or abilities and those who do not. There are two alternative responses to the *epistemic critique of democracy.* First, in a nonepistemic approach to democracy, one devises an account of the value of democracy that does not depend on its having any particular epistemic virtues—any particular tendency to make good decisions. Second, in an epistemic approach to democracy, one accepts the epistemic demands and argues that democracy *can* meet them.

Some theories argue that (properly arranged) democratic procedures are epistemically the best methods possible. The losing voter is expected to take the outcome of the vote as her best evidence on the question and, presumably, to change her mind (a classic example of such a thesis being Rousseau's theory). An alternative view, espoused by David Estlund, denies that greater wisdom gives anyone the right to rule unless their expertise is itself beyond reasonable denial.

Many advocates of democracy have preferred a *nonepistemic* approach. One reason is that the idea of a fair political procedure by which the people rule themselves might account for our democratic convictions even without any appeal to good decisions. Casting democracy as a right to self-rule, however, threatens to neglect the obvious fact that voters decide the rules for others as well as for themselves. Emphasizing fair procedure rather than self-rule prompts a distinct worry. The thin idea of an equal procedure does not justify voting procedures for choosing policies rather than random procedures

such as flipping a coin. If voting is to be preferred to random methods, it is evidently not on grounds of fair procedure alone and might even be on unstated epistemic grounds.

An important mathematical result that derives from Marquis de Condorcet (in his *Essay on the Application of Analysis to the Probability of Majority Decisions,* 1775) shows that a group of voters, each of whom is better than random, using majority rule, can perform better even than the best of the individual voters. Performance improves as the group gets larger, approaching infallibility. Applying this "jury theorem" to the case of democratic voting is problematic for at least the following three reasons: (1) the mathematical results are far less impressive if the choice is between more than two alternatives; (2) voters who are influenced by other voters in certain ways do not allow for the full mathematical effect, and so deep structures of influence can dampen the results significantly; and (3) voters who, as most humans do, have various prejudices, blind spots, or rational defects might be worse than random, in which case the jury theorem drives the group's competence just as impressively toward certain error.

The jury theorem approach doesn't rely on any epistemic benefits of communication, as other epistemic approaches do. In some contexts, interpersonal communication clearly improves a group's epistemic competence, and public deliberation about political decisions may have some of these epistemically beneficial features. Critics of "deliberative democracy" (a family of views that only sometimes takes this epistemic form), however, argue that interpersonal deliberation is often distorted by power dynamics, by tendencies to follow the herd, and so on.

On some views, the standards of correctness for decisions are independent of the deliberative process, just as a sum of two numbers is independent of the method used to find it. On other views, including several "pragmatist" epistemic approaches to democracy influenced by Dewey and others, a just or correct political decision *is* one that was (or would have been) produced by properly arranged public political deliberation. A closely related family of views also holds that the standards are constituted by actual or hypothetical deliberative processes, not (as in the pragmatist variant) processes of inquiry but processes of individual pursuit of interests in a morally appropriate framework.

Epistemic approaches do not assume that actually existing democracies make good decisions, or even that greatly improved arrangements would lead to predominantly good decisions. What makes them count as epistemic approaches is their holding that an adequate normative theory of democracy's legitimacy, authority, or justification depends partly on some tendency of (some) democratic arrangements to make good political decisions by aiming to do so.

David Estlund

See also Collective Rationality; Communicative Action Theory; Decision Theory; Judgment Aggregation and the Discursive Dilemma; Pragmatism; Public Reason and Justification; Social Epistemology

Further Readings

Christiano, T. (2008). *The constitution of equality: Democratic authority and its limits.* Oxford, England: Oxford University Press.

Cohen, J. (1986). An epistemic conception of democracy. *Ethics, 97*(1), 26–38.

Coleman, J. L., & Ferejohn, J. (1986). Democracy and social choice. *Ethics, 97*(1), 6–25.

Estlund, D. (2008). *Democratic authority: A philosophical framework*. Princeton, NJ: Princeton University Press.

Grofman, B., & Feld, S. (1988). Rousseau's general will: A Condorcetian perspective. *American Political Science Review, 82*(2), 567–576.

Habermas, J. (1996). Appendix 1. Popular sovereignty as procedure. In J. Habermas (Ed.), *Between facts and norms: Contributions to a discourse theory of law and democracy* (W. Rehg, Trans.; pp. 463–490). Cambridge: MIT Press.

List, C., & Goodin, R. E. (2001). Epistemic democracy: Generalizing the Condorcet jury theorem. *Journal of Political Philosophy, 9*(3), 277–306.

Meiklejohn, A. (1948). *Free speech and its relationship to self-government*. New York, NY: Harper.

Nino, C. (1996). *The constitution of deliberative democracy.* New Haven, CT: Yale University Press.

Rousseau, J. J. (2008). *The social contract*. New York, NY: Cosimo. (Original work published 1762)

Sunstein, C. (2006). *Infotopia: How many minds produce knowledge.* Oxford, England: Oxford University Press.

Talisse, R. (2009). *Democracy and moral conflict.* Cambridge, England: Cambridge University Press.

Waldron, J. (1999). *Law and disagreement.* Oxford, England: Oxford University Press.

EPISTEMOLOGY

Epistemology is the philosophical subdiscipline that studies the evaluative dimensions of cognition, their metaphysical bases, and, increasingly nowadays, the language we use to ascribe cognitive achievements. The nature and scope of knowledge is the central focus of epistemology.

Knowledge and Assertion

Assertion is a speech act whereby we communicate to our audience that a certain proposition is true, and it is the main way we communicate information to one another. Given that we rely so pervasively on assertions, it is natural to wonder what the standards ought to be for our assertions and whether we as assertors live up to them? Research into the epistemic norms of assertion is one of the most exciting and rapidly growing fields within philosophy. Within this field, the leading proposal is that knowledge is the norm of assertion, a view known as *the knowledge account of assertion.*

The knowledge account of assertion says that knowledge sets the standard for permissible assertion: You may assert *P* only if you know that *P* is true. The primary evidence for the knowledge account is an explanatory inference from linguistic data, in particular conversational patterns surrounding the give-and-take of assertion. For example, typically, when you make an assertion, even when the content of your assertion has nothing to do with you or what you know, it is appropriate to ask you "How do you know that?" The knowledge account explains the appropriateness of the question as follows: In making the assertion, you represent yourself as having the authority to do so, and knowledge is what gives you the authority, rendering it sensible to ask how you know that. Similarly, if someone makes an assertion, the response "You don't know that" is taken as an outright rejection of his or her authority to make the assertion, which the knowledge account easily explains. Also, when someone asks you a question, even when the question has nothing to do with you or what you know, it is typically appropriate for you to respond, "I don't know." The knowledge account explains why it is appropriate: By saying "I don't know," you inform the questioner that you lack the authority to answer the question.

Those attracted to the knowledge account of assertion are often attracted to related views about practical reasoning, action, and belief, namely, that knowledge is the norm of these too. These other three accounts are also supported by linguistic data and patterns of appraisal, though the overall case for them is generally regarded as weaker than the case for the knowledge account of assertion.

Furthermore, investigation into the epistemic norms of assertion, reasoning, action, and belief presents opportunities for collaboration among epistemologists and social and cognitive scientists.

Contextualism

Skepticism perennially fascinates philosophers. It is almost unanimously accepted that skepticism is false, so one main question is "Why do skeptical challenges nevertheless seem so formidable?" For example, it usually seems obvious that when you say, for instance, "I know that I have hands," you speak truthfully. But do you know that you are not a handless brain in a vat (BIV) being perfectly deceived into thinking that you have hands? Most people hesitate at this question, and many go so far as to say, "No, I don't know that I'm not a handless BIV." Yet if you know that you have *hands,* then you seem to be in an ideal position to simply deduce, and thereby know, that you are not a *handless* BIV. So why are we tempted to say "I don't know" in such a situation?

Setting aside extravagant skeptical scenarios featuring BIVs, we see a similar effect when speakers move from an ordinary context to a high-stakes context. If a fellow traveler in an airport looks up from his magazine and nonchalantly asks, "I'm bored. Do you know whether this flight we're about to board goes directly to Chicago?" You might check the board, see that it says "Direct flight to Chicago" and respond, "Yes, I happen to know that it does." But suppose that instead of being asked by a nonchalant fellow traveler, you are asked the same question by an organ courier transporting an organ to a patient in Chicago, which will spoil unless she takes a direct flight to Chicago. Now, even if the board says "Direct flight to Chicago," it would be natural for you to respond by saying, "No, sorry, I don't know. Maybe you should go ask the captain." Again, why are we tempted to say "I don't know" in such a situation?

In recent years, some epistemologists have proposed a *semantic* solution to the apparent conflict in what we are willing to say about knowledge in ordinary contexts versus what we say when confronted with skeptical hypotheses or high-stakes decisions. *Epistemic contextualism* is the view that the cognitive verb *knows,* as it features in propositional knowledge ascriptions such as "I know that *P*" and "You don't know that *P,*" is a context-sensitive term. The truth conditions of "I know that this is a direct flight" are different in an ordinary context from those in a high-stakes context; in an ordinary context, less evidence is required for you to truly say "I know that this is a direct flight" or "I have hands" than is required in a context where the stakes are high or where serious skeptical hypotheses have been raised. Thus, the contextualist can maintain *both* that we typically speak truthfully when we say "I know this" or "I know that" *and* that the skeptic speaks truthfully when he says "You don't know this" or "You don't know that." Despite appearances, what the skeptic says does not contradict what we ordinarily say, just as I do not contradict you when I say (in Vancouver) "It's raining here" and you say (in Paris) "It's not raining here."

Contextualists disagree over the correct semantic model for "knows." Some contextualists say that "knows" is an indexical expression, similar to the pronoun *I* or the adverb *yesterday.* Others liken it to gradable adjectives, such as *tall* and *flat.* But both of those semantic models have been severely criticized. A more recent suggestion is that knowledge ascriptions are context-sensitive not because "knows" is context-sensitive but because *in general* the propositional content of speech acts is context-sensitive and knowledge ascriptions are no exception. The semantics of "knows" and knowledge ascriptions is ripe for interdisciplinary collaboration between epistemologists and linguists. Research into the psychology of judgments made in high-stakes contexts or in the context of skeptical challenges could also be highly relevant to sorting out the linguistic data. And anthropological data on the linguistic behavior of humans in other cultures could highlight further important data and theoretical possibilities.

The Value of Knowledge

It is a virtual platitude that knowledge is intellectually better than mere true belief. Going all the way back to Plato's dialogue *Meno,* philosophers have asked why this is so. One deceptively simple answer is that knowledge is better because it has greater practical value than true belief. But Plato ruled this out when he first raised the question, noting, for example, that merely truly believing this is the road to Larissa will get you to Larissa just as well as knowing that this is the road to Larissa will. So not only does it seem correct that knowledge is better than mere true belief, it also seems that knowledge's added value is not merely practical. Some epistemologists contend furthermore that knowledge is better than mere true belief not only in degree but also in kind: Knowledge has a kind of intellectual value that mere true belief lacks.

It is widely accepted that an adequate theory of knowledge must explain knowledge's special value, though this is not entirely uncontroversial. There is a consensus, however, that it is at least a good thing for a theory to explain knowledge's value.

Several theories have been proposed to explain knowledge's value. On one view, knowledge is valuable because it is the most general factive propositional attitude. Factive propositional attitudes are attitudes that you can have only toward a true proposition; this class of attitudes includes *remembering that P, perceiving that P,* and *being aware that P.* To say that knowledge is the most general factive propositional attitude is to say that the following is a necessarily true generalization: If you bear any factive propositional attitude toward the proposition *P,* then you know that *P.* We value a match between mind and world, and knowing is the most general attitude in which the mind must match the world, which according to this view explains why knowledge is specially valuable.

A different proposal begins by pointing out that while it is, of course, good to have good things, it is even better to *merit* or *earn* the good things you have. For example, other things being equal, it is better to earn your fortune than to win it through the lottery. Put otherwise, it is better to succeed through skill and effort than through luck. So if knowledge were true belief for which you earn credit or if knowledge were true belief manifesting intellectual skill, then that would explain why knowledge is better than mere true belief. For then the superiority of knowledge would just be a special case of the superiority of earning something good or achieving it through skill, as opposed to just getting lucky. This

explanation of knowledge's value is distinctive of an increasingly popular approach in epistemology called *virtue epistemology.*

The Gettier Problem

Perhaps the most important development in epistemology over the past 50 years was the research program ignited by Edmund Gettier's three-page article published in 1963, "Is Justified True Belief Knowledge?" Many philosophers at the time thought that knowledge could be defined as justified true belief (JTB for short). Gettier presented what most philosophers took to be decisive counterexamples to the JTB account of knowledge. These examples were taken to show that JTB is not sufficient for knowledge—that something more than JTB is required.

Gettier cases follow a recipe. Start with a belief sufficiently justified to meet the justification requirement on knowledge. Then, add an element of bad luck that would normally prevent the justified belief from being true. Last, add a dose of good luck that counteracts the bad luck, so that the belief ends up true anyhow. It has proven difficult to explain why this "double-luck" structure prevents knowledge.

Philosophers have proposed dozens of hypotheses to handle these "Gettier cases," as they came to be called. It is widely assumed that an acceptable theory of knowledge must explain what goes wrong in Gettier cases. No-false-basis theorists argued that knowledge is JTB, which is not essentially based on a falsehood, but Gettier cases all involve beliefs essentially based on a falsehood. Defeasibility theorists argued that knowledge is JTB, for which there is no further fact such that it would defeat the subject's justification if she learned of it, but Gettier cases all involve facts that defeat justification. Reliabilists argued that knowledge is reliably produced true belief, but Gettier cases are examples of unreliably produced belief. Causal theorists argued that knowledge is true belief nondeviantly caused by the fact believed to be true, but the right kind of causal link is missing in Gettier cases. Counterfactual theorists argued that knowledge is true belief that has the right counterfactual relationship to the truth believed (the believer would get it right across a relevant class of counterfactual scenarios), but in Gettier cases, the relevant counterfactual relation is missing. Virtue epistemologists argue that knowledge is true belief manifesting intellectual virtue, but in Gettier cases, the subject's true belief fails to manifest her virtue.

None of these solutions has been widely accepted. At least a few epistemologists claim that the entire literature on the Gettier problem is mistaken because it falsely presupposes that the subject in a Gettier case lacks knowledge. On one such view, knowledge is merely true belief, though this view is almost universally rejected.

Epistemology and Experimental Philosophy

Experimental philosophers have made a very interesting and potentially significant sociological observation: People from East Asian cultures have a greater tendency than people from Western cultures to judge that the subject in a Gettier case has knowledge. More generally, results from experimental philosophy demonstrate that differences in gender, age, socioeconomic status, and culture are often associated with statistically significant differences in intuitive judgments about thought experiments of the sort Gettier used and that philosophers routinely rely on in the course of debate and inquiry. Uncovering these often very surprising differences and understanding their significance is an area of extremely fruitful and rapidly increasing collaboration among philosophers and psychologists.

John Turri

See also Common Knowledge; Epistemology of Mass Collaboration; Ethno-Epistemology; Experimental Philosophy; Feminist Epistemology; Normativity; Social Epistemology; Speech Acts; Virtue Epistemology

Further Readings

Dancy, J., Sosa, E., & Steup, M. (Eds.). (2011). *A companion to epistemology.* Malden, MA: Wiley-Blackwell.

DeRose, K. (2009). *The case for contextualism.* Oxford, England: Oxford University Press.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis, 23,* 121–123.

Haddock, A., Millar, A., & Pritchard, D. (Eds.). (2009). *Epistemic value.* Oxford, England: Oxford University Press.

Hawthorne, J., & Stanley, J. (2008). Knowledge and action. *Journal of Philosophy, 105,* 571–590.

Knobe, J., & Nichols, S. (Eds.). (2008). *Experimental philosophy.* Oxford, England: Oxford University Press.

Sosa, E. (2007–2009). *Apt belief and reflective knowledge* (Vols. 1–2). Oxford, England: Oxford University Press.
Sosa, E., Kim, J., Fantl, J., & McGrath, M. (Eds.). (2008). *Epistemology: An anthology.* Malden, MA: Blackwell.
Stainton, R. (2010). Contextualism in epistemology and the context-sensitivity of "knows." In J. Campbell, M. O'Rourke, & H. Silverstein (Eds.), *Knowledge and skepticism* (pp. 137–163). Cambridge: MIT Press.
Steup, M., Turri, J., & Sosa, E. (2012). *Contemporary debates in epistemology* (2nd ed.). Malden, MA: Wiley-Blackwell.
Turri, J. (2011). The express knowledge account of assertion. *Australasian Journal of Philosophy, 89,* 37–45.
Williamson, T. (2000). *Knowledge and its limits.* Oxford, England: Oxford University Press.
Zagzebski, L. (2009). *On epistemology.* Belmont, CA: Wadsworth.

Epistemology of Mass Collaboration

People usually work together to get things done. In particular, they often team up to find things out. New information technologies now allow large numbers of people to collaborate in order to discover and to disseminate knowledge. A notable example is Wikipedia, the "free online encyclopedia that anyone can edit."

The *epistemology of mass collaboration* is the study of when, and why, such large collaborative efforts can be effective at producing knowledge. In particular, when, and why, is the information produced by large groups of people reliable?

This entry looks at how this topic relates to epistemology in general, provides some examples of mass collaboration, and discusses what is known about the reliability of large groups.

Connections With Epistemology

Roughly speaking, epistemology is the study of (a) what knowledge is; (b) what sorts of things, if any, people can know; and (c) how people can come to know these sorts of things. Epistemology has traditionally focused on how people can come to know things working on their own (e.g., using perception or reasoning). However, most people acquire most of their knowledge from other people. The prototypical case is where one person tells another person something. But the ultimate source of our information about the world is often a group of people who have worked together to make a discovery. Thus, the epistemology of mass collaboration falls within the domain of *social epistemology* and the *epistemology of testimony.*

Examples of Mass Collaboration

It is fairly common for scientists to work together in teams to discover facts about the physical world. Although such scientific research teams are usually fairly small, they have steadily been getting larger. In fact, papers in high-energy physics with hundreds of coauthors are not uncommon.

New information technologies now allow even larger numbers of people, separated by very large distances, to work together on a single project. For instance, thousands of programmers from around the world have collaborated on open-source software projects, such as the GNU/Linux operating system. *Yahoo! Answers* is a social networking project that allows people to get answers to just about any question by leveraging the collective wisdom of many other Internet users. The *Iowa Electronic Markets* allow people to buy and sell contracts about future events, and the value of each contract can be used to predict whether the event in question will actually occur.

There are two main ways to leverage large groups to discover and to disseminate knowledge: (1) the project can involve actual collaboration, as when many people make contributions to an entry in Wikipedia and/or (2) the project can involve simply aggregating the separate opinions of the members of the group, as with prediction markets.

It seems plausible that a large group of experts would be more reliable than a single expert. For instance, scientific research teams are usually made up of people whose areas of expertise complement each other. But many large collaborative projects are open to anyone who is interested in participating. So, for example, there is no guarantee that the person writing or editing the Wikipedia article on bioethics has any training or expertise in bioethics. This lack of expertise has led to worries about the reliability of such large collaborative projects.

The Wisdom of Crowds

Researchers such as James Surowiecki and Scott Page have found that *sufficiently large* groups of nonexperts can be surprisingly reliable. For instance, on the television show *Who Wants to Be a Millionaire?* when a contestant is stumped by a question, she can poll the studio audience or phone a friend for help. It turns out that consulting the collective wisdom of the audience is a much more reliable "lifeline" than consulting your smartest friend. This phenomenon is typically referred to as the *wisdom of crowds*.

A notable result in game theory, known as the *Condorcet Jury Theorem,* helps explain this phenomenon (from Marquis de Condorcet's *Essay on the Application of Analysis to the Probability of Majority Decisions,* 1775). Suppose that we ask the members of a large group to vote on a factual yes-or-no question. According to the theorem, even if the individual members are not all that reliable on this question, the group as a whole is very reliable. Moreover, the bigger the group is, the more reliable it is. In fact, if the group is big enough, the majority vote is almost certainly correct.

This phenomenon also seems to apply to social networking projects such as Wikipedia. A study published in the journal *Nature* found that Wikipedia is only slightly less reliable on scientific topics than *Encyclopedia Britannica*. A partial explanation for Wikipedia's reliability is suggested by the saying in the software industry that "given enough eyeballs, all bugs are shallow." Any errors in Wikipedia can be corrected very quickly because any of the millions of people reading the encyclopedia can make corrections.

The Properties of Wise Groups

But not just any large group of people is going to be reliable on just any question. For instance, a large group of nonexperts could *introduce* errors into Wikipedia just as easily as they could correct errors. To be wise, a large group must have certain properties.

Indeed, the Condorcet Jury Theorem does not simply assume that there are a lot of voters. It also assumes each voter is at least slightly more likely to give a correct answer than an incorrect answer. In addition, the theorem assumes that their votes are *independent*. In other words, there is no correlation between the probability that one voter will vote a certain way and the probability that another voter will vote that same way.

Surowiecki hypothesizes that wise groups must be diverse, independent, and decentralized. Independence is a very strong requirement. Fortuitously, as Page has discovered, it is actually diversity rather than independence that makes a group wise. Independence is just a good way to ensure that there is a reasonable degree of diversity in a group.

Don Fallis

See also Collective Agents; Epistemic Approaches to Democracy; Epistemology; Judgment Aggregation and the Discursive Dilemma; Philosophy of Expertise; Social Epistemology

Further Readings

Fallis, D., & Sanger, L. (Eds.). (2009). The epistemology of mass collaboration [Special issue]. *Episteme, 6*(1).

Page, S. E. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton University Press.

Surowiecki, J. (2004). *The wisdom of crowds*. New York, NY: Doubleday.

Equilibrium in Economics and Game Theory

At the core of modern economics lies the notion of *equilibrium*. The possibility of identifying how the actions of different agents eventually define a resting point provides the economic analysis with a powerful instrument to foresee some of the consequences of different economic policies. At the same time, the notion of equilibrium is central to game theory.

This entry discusses the most widely used definitions of equilibrium and their relation to game theory, applied especially to the domain of economic behavior. Critical philosophical issues are shown to be raised by such an equilibrium methodology in economics.

The Walrasian Equilibrium

Adam Smith's notion of the invisible hand is perhaps the first description of an economic equilibrium. One century later, the French mathematical economist Léon Walras provided a rigorous definition:

The equilibrium is achieved when the plans of both sellers and buyers are compatible with each other in a precise, mathematical way. However, although Walras's theory gives a compelling characterization of the equilibrium point, it fails to explain how the economy gets there in the first place.

Walras himself postulated a "*tatonnement* process" to solve the problem. If there is an excess supply, the price falls; if there is an excess demand, it rises. Theoretically, this process is simple, but it is not entirely satisfactory because it relies on an ad hoc behavioral hypothesis and is not specific enough to guarantee that, out of equilibrium, the economic system will converge to it. The process is not without merit, either, as the ad hoc hypotheses are not too demanding, and the process does have explanatory power in many instances and is good enough if one does not pursue mathematical purity.

When confronted with reality, the *tatonnement* process can also be criticized because a market often seems to be at equilibrium and yet shows an excess supply or demand. Involuntary unemployment, for instance, may be persistent. This opens the door to more complete models of behavior and to alternative definitions of equilibrium and to dynamics.

The Value of a Zero-Sum Game

Games are mathematical models to study strategic decision making. Zero-sum games are games in which the interaction among players does not create anything, like in a poker game or in the game of cutting a cake. In this class of games, and for the case of two players, the mathematician John von Neumann defined the concept of value, a payoff V with the property that Player 1 can guarantee himself V and Player 2 can guarantee herself $\neg V$. Then von Neumann proved the minimax theorem, which states that every two-person zero-sum game has a value. Furthermore, there is a nice behavioral procedure that leads to the value: If players play their maxmin strategies, they will receive the value. In a maxmin strategy, a player considers first the worst-case scenario after each option and then chooses the strategy that has the best of the worst outcomes. This is the best we can get—a definition of an equilibrium and a theory about how to get there. Unfortunately, this cannot be generalized to all games.

The Nash Equilibrium

When playing a non–zero-sum game, things get more complicated, and von Neumann's approach does not work. The mathematician John Nash proposed instead his theory: An equilibrium is defined as a choice of strategies in which no player has an incentive to change her or his behavior given the choice made by the others (a "Nash equilibrium").

Games are more general models than markets, and in many respects, the Nash equilibrium is a more general concept than Walras's. For instance, models to study oligopolistic behavior can be analyzed using the concept of the Nash equilibrium, which converges to the Walrasian equilibrium as the number of agents increases. It is thus no surprise that both share some methodological problems; in particular, the rationality assumptions in game theory are not enough to define a process leading toward a Nash equilibrium, thus repeating the situation found with the Walrasian equilibrium.

Dynamics

There are two main reasons to analyze the dynamics of behavior. First, a dynamic process that converges to the equilibrium may respond to the methodological criticism mentioned before. Second, the equilibrium analysis may require too much rational and analytical capability from the agents. A dynamics based on a learning rule, an adaptive process, an imitation reflex, a contagion mechanism, and so on, may explain better the observed reality.

In the end, the dynamic process may or may not converge to a strategy for each agent. If it does, the point of convergence is a kind of equilibrium, which typically, but not necessarily, is a Nash equilibrium. If it does not converge, the process can cycle or may show an erratic behavior.

The Use of the Equilibrium Methodology

The General Equilibrium Theory studies the conditions under which markets are efficient. These conditions serve two purposes: They guarantee, first, that a competitive equilibrium exists and, second, that the competitive equilibrium is efficient. This distinction is important in order to understand the scope of the amendments that are introduced in the theory to cope with other conditions. Next, we discuss three cases.

The absence of free entry, for instance, opens the door to a model of oligopolistic behavior where one can compute both the Walrasian and the Nash equilibria, which give different outcomes. The choice of the Nash equilibrium is made after the observation that the strategic and not the price-taking behavior is the more natural behavioral assumption in this setting.

The equilibrium analysis shows that markets are inefficient if externalities are present (pollution is just one example) and provides a diagnosis of the problem that may help design a correcting mechanism. Depending on the context, a bilateral negotiation after assigning property rights, a tax, or an emissions market are examples of such mechanisms. Sometimes the model will be solved by using the competitive equilibrium and some other times by using the Nash equilibrium.

In economics, as in game theory, rationality means not only that preferences and behavior must be consistent but also that agents can understand and analyze the situation as well as the modeler can. However, collecting all the information and analyzing all the data may be too demanding, and models that use a dynamic process of learning, imitation, or evolution, to name a few, may cope with this problem.

Problems With the Equilibrium Methodology

Realism

Critics of the equilibrium theory argue that the economy is never at rest and that the equilibrium is never attained. Two solutions can be found when confronted with this problem. First, if the state is not far from the equilibrium, the equilibrium methodology could be defended as a reasonable approximation to the observed phenomena. Second, if the approximation is not good enough, one probably needs to model behavior as a dynamics.

Existence

The determination of the conditions for the existence of an equilibrium will be important not only for the acceptance of the theoretical model but also as an indication of what to look for in real life to address the question of whether and when an economy can be said to be close to an equilibrium.

Multiplicity

A model that allows for multiple equilibria shows an undesired level of uncertainty. A better model may need to incorporate more details to have more predictable power. Alternatively, a refinement of the equilibrium concept may be enough to fix the problem. Which one is the best option cannot be determined a priori.

Stability

One particularly important refinement discards the unstable equilibria, as the real economy can hardly be in such a situation. The modeler needs to be able to tell them apart and use only the stable equilibria as her tool of analysis.

José Luis Ferreira

See also Evolutionary Game Theory and Sociality; Feedback Mechanisms and Self-Regulatory Processes in the Social Sciences; Game-Theoretic Modeling; Markets and Economic Theory; Mathematical Models, Use in the Social Sciences; Reflective Equilibrium

Further Readings

Aumann, R. (1985). What is game theory trying to accomplish? In K. Arrow & S. Honkapohja (Eds.), *Frontiers of economics* (pp. 28–76). Oxford, England: Blackwell.

Debreu, G. (1959). *Theory of value, an axiomatic analysis of economic equilibrium*. New Haven, CT: Yale University Press.

Mosini, V. (Ed.). (2007). *Equilibrium in economics: Its scope and limits*. London, England: Routledge.

Nash, J. (1996). *Essays on game theory*. Cheltenham, England: Edward Elgar.

Schmidt, C. (Ed.). (2002). *Game theory and economic analysis: A quiet revolution in economics*. London, England: Routledge.

Van Daal, J., & Jolnik, A. (1993). *The equilibrium economics of Léon Walras*. London, England: Routledge.

Varian, H. R. (1992). *Microeconomic analysis* (3rd ed.). New York, NY: W. W. Norton.

Von Neumann, J., & Morgenstern, O. (2004). *Theory of games and economic behavior* (60th anniversary ed., with A. Rubinstein & H. W. Kuhn). Princeton, NJ: Princeton University Press. (Original work published 1944)

Essentialism

Rather than one view that goes by the name *essentialism,* there are many. This entry clarifies and emphasizes important differences and other relations between some varieties of essentialism.

But it is useful to begin with what most essentialisms have in common. They are views about *belonging*. Most of them further specify the general idea that for some entity to belong to a particular category or individual, it must have certain properties. Those properties are *essential properties* for belonging to the category or individual in question. For example, essentialism about the biological category "platypus" might say that an organism is a platypus *only if* that organism has a certain cluster of genes and a particular ancestry, where these genes and ancestry are essential properties for being a platypus. Essentialism about the March 2011 dissolution of the 40th Canadian Parliament (an individual event in Canadian politics) might say that any event belonged to or was a part of that dissolution *just in case* it was one of the important proximate causes of that dissolution. Additionally, most essentialisms imply that the properties essential for belonging to some category or individual together form the *essence* of (or essence of belonging to) that category or individual. If having a particular cluster of genes and a certain ancestry are the *only* essential properties of being a platypus, then together they form the platypus essence. More generally, an essential property for belonging to *X* is *necessary* for belonging to *X;* having the essence for belonging to *X* is, in ordinary environments, *sufficient*.

To see how varieties of essentialism elaborate this basic view in different ways, this entry first discusses *philosophical* essentialisms, including those in metaphysics and philosophy of science that are relevant to thinking about the nature of social science categories such as "economic individual," "urban city," "Black person," and "gay man." The entry then more briefly discusses *psychological* essentialisms concerning folk beliefs about such categories. The discussions reveal how several essentialist views connect with other issues, including categorization of kinds, induction, scientific realism, explanation, social constructivism, reductionism, the psychology of concepts, and social policy.

Philosophical Essentialisms

Of the many philosophical essentialist views and issues, this entry discusses metaphysical and then scientific ones.

Philosophy: Metaphysical Essentialist Views and Issues

Ancient Greek philosophers such as Aristotle were the first thinkers known to develop essentialist views, typically, to address metaphysical problems. These include explaining how anything can be generated out of other things and how a thing can persist through some changes but not others. One might say that a particular platypus can survive the loss of its tail because having a tail is not an essential property for being a platypus.

The popularity of metaphysical essentialisms has fluctuated dramatically since Aristotle's time. Working in the 1970s on the issue of linguistic reference (e.g., To what do the terms *Richard Nixon* and *gold* respectively refer?), the American philosophers Saul Kripke and Hilary Putnam initiated the present resurgence in metaphysical essentialisms. Some of these abstract from any particular sorts of entities to focus on entities in general (including objects, processes, events, groups, nations, etc.). These views often articulate theories about what properties entities have *necessarily.* A spectrum of such views range from the claim that any entity has all its properties necessarily, as the German philosopher Gottfried Leibniz claimed in the 17th century, to the view that the only properties that any entities have necessarily are trivial ones, such as the property of being either red or not red.

As metaphysicians have made their essentialist views increasingly responsive to work in other fields, general essentialisms have fragmented into more specific views about limited ranges of entities, for instance, linguistic, biological, or social entities. Nevertheless, some general metaphysical issues arise across several of these narrower views. One issue concerns the *ontological* categories with which essential properties associate. For example, take the view that some of an entity's essential properties are those without which it would not exist. Such properties individuate (set apart) the fundamental kinds to which entities belong, where "fundamental kind" names an ontological category. The essences formed by these essential properties are *kind essences*.

Another sort of kind essence is one that individuates a nonfundamental kind, one that an entity can pass in and out of without perishing. The essential properties constituting such an essence are required for belonging to the nonfundamental kind but not necessarily for existing. If "republic" names a kind, it is of this sort; Australia doesn't currently possess the essential properties for being a republic, but it might one day. Some of the problems that scientists and philosophers of science have perceived for essentialism tacitly presuppose that essentialism is only about fundamental kinds; but many kinds of interest in science are nonfundamental, and so such problems do not afflict them.

All kind essences, whether associated with fundamental kinds or otherwise, are ones that, in principle, more than one entity can have. Kinds can have more than one member. *Individual essences,* in contrast, are made of properties essential to and had by only one entity. If Australia has an individual essence, it probably involves the unique way it originated.

The ontological issue relates to *identity.* Essences that individuate fundamental kinds determine the metaphysical identities of the members of those kinds. For instance, were *platypus* a fundamental kind, then any organism with the platypus essence would fundamentally *be* a platypus; that essence would fix the organism's identity so that it could not cease to be a platypus without expiring altogether. Interestingly, some plausible interpretations of biology suggest that the species that evolutionary theory recognizes are not fundamental (kinds or otherwise), allowing that any organism could survive a change in species and even belong to no species or to more than one. This does not imply that *every* essentialism about these species is hopeless, because it leaves it open that each species that evolutionary theory recognizes is individuated by essential properties of the sort associated with nonfundamental kinds, such as "republic." Likewise, essentialists with respect to social science categories such as "free market," "woman," or "gay man" need not claim that the essential properties they recognize determine identity.

Identity issues underlie views concerning the *locality* of essential properties. Intrinsicalism is a common though seldom defended presumption about the locality of essential properties; it says that any essential properties must be *intrinsic* properties of their bearers. A subject's intrinsic properties are realized by that subject's internal features (e.g., some muscle internal to your chest realizes your property of having a heart). One motivation for intrinsicalism applies *only* to essentialism about fundamental kinds. Consider the following: Were any of the essential properties that individuate these kinds not intrinsic, they would be *extrinsic.* A subject's extrinsic properties are realized at least in part by features external to her, such as being related to other entities or processes in particular ways. But if such extrinsic properties are among those essential for belonging to a fundamental kind, then, absurdly, members of those kinds could perish due to changes in their extrinsic conditions and without any internal changes to themselves. In contrast, because the essential properties associated with nonfundamental kinds do not determine identity, they *can* be extrinsic without generating such absurdities. More generally, several authors have argued that there is no barrier to extrinsic property essentialism about nonfundamental kinds and that we often have good theoretical reasons to recognize these. This allows such forms of essentialism to agree with the common claim that membership in many social categories, such as "free agent" or "Canadian," is (partly or wholly) extrinsically or relationally determined.

Several other metaphysical nuances in contemporary essentialist views belie traditional understandings of essentialisms. Essentialism about kinds is often said to imply *fixity* or *immutability* about hierarchies or networks of kinds, for example, that all kinds of chemical elements form an unchanging and static set that the periodical table represents. But many essentialisms are compatible with dynamic networks of kinds, where some kinds are generated out of others. Indeed, the chemical elements probably arose in this way. Some authors worry that this ensures that the boundaries of these essentially determined kinds are *vague,* making nonarbitrary identification of them impossible. Authors such as Elliott Sober reply that there are reasons to think that a kind's having a vague essence and vague boundaries is compatible with its being determinate and nonarbitrary.

Perhaps the most startling check on traditional metaphysical presumptions about essentialism concerns *necessity.* On the increasingly popular homeostatic property cluster (HPC) view of some kinds, no single one of the properties that helps individuate an HPC kind need be necessary for kind membership;

rather, some subset of these properties is sufficient in each case. If "Irish person" names an ethnic kind, for example, it is probably one of these. There need not be one property that all Irish people share but rather a cluster of individuative properties of which each has some subset, with different Irish people having different subsets. (Some other people are neither determinately Irish nor determinately not Irish.) This descendant of Ludwig Wittgenstein's notion of "family resemblance" would be a nonessentialist view were it not for the possibility that the individuative cluster *is* necessary rather than any single property in it. On this possibility, although there is no single property that each Irish person must have, to be Irish a person *must* have some subset or other of the properties in the cluster. Such views have been made consistent with Kripke's and Putnam's work on linguistic reference and define a new form of essentialism that permits the prodigious variation within kinds that dooms many traditional essentialist accounts of those kinds.

Philosophy: Scientific Essentialist Views and Issues

Metaphysical and scientific essentialisms overlap. Work in various sciences and philosophy of science motivates some of the nuanced metaphysical positions described above. In the other direction, refinements in metaphysics have made essentialist views more applicable in some scientific domains than they previously were. Nevertheless, teasing out scientific essentialist views and issues from metaphysical ones clarifies issues both in *general philosophy of science* and in philosophies of *particular sciences*. Let us take these in turn.

In addition to renewing interest in essentialism among metaphysicians, Saul Kripke and Hilary Putnam helped rejuvenate the Aristotelian idea that essentialism is important to empirically minded philosophy of science. They did this partly by convincing many researchers that some scientific inquiry consists in an empirical search for and a posteriori discovery of kind essences. For example, chemistry has involved not only searches for chemical causes of certain phenomena but also determining what particular chemical elements and compounds *are*. Their results often seem to tie these kinds to *microstructural* essences: Having 79 protons is the essence of being a gold atom, and being composed exclusively of H_2O molecules is the essence of being (pure) water. Kripke's and Putnam's intuitively driven thought experiments have had a greater role in inspiring an essential interpretation of some scientific inquiry than their semantic theses.

Brian Ellis has developed an essentialist interpretation of some scientific inquiry, in a view he calls *new scientific essentialism*. Although he hesitates to apply the view to the social sciences, we will see why others are cautiously optimistic.

Ellis claims that his scientific essentialism best fits the facts of inquiry and discovery in much of physics and chemistry and that it offers the best philosophical analysis of the laws of nature discovered in those disciplines. Roughly, he proposes that laws of nature are grounded in the microstructural essences of the kinds over which those laws range. The laws are exceptionless *because* the members of the kinds over which the laws range all possess the essences that make the laws true. According to Ellis, those essences are metaphysically necessary and sufficient for kind membership.

On Ellis's view, essential properties of scientific interest must be an important part of the *explanation* of characteristic behaviors of the entities that have them. Having 79 protons, for instance, is an important part of the explanation of gold's melting behaviors and interactions with other elements. This explanatory salience of essential properties is supposed to make them relevant to the general philosophy of science issue of *induction* too. The explanatory salience of the essential properties ensures the predictive reliability of generalizations that range over the kinds individuated by those properties.

The new scientific essentialism connects with the further issue of the proper aims of science. It supports the traditional view that some sciences do and should aim to construct classifications comprising categories that represent kinds over which generalizations range. The classifications are theories about natural order. These improve as they more accurately represent kinds, laws, or generalizations and the relations between these. Whereas empiricists about the aims of science often transform this view into an instrumentalist or anti-realist one that nowhere appeals to microstructural essences, Ellis argues that his foundational appeal to such essences develops the view into a version of scientific realism. He claims that science discovers these real essences and discovers that they are essences and that they are intrinsic causal powers that members

of corresponding kinds have determinately without variation. The kind distinctions they underwrite are then nature's distinctions, not ours: real, absolute, and categorical.

Philosophy of the Social Sciences

The reality of essential distinctions and the naturalness of kinds are hotly contested issues when one moves from general philosophy of science to the philosophy of particular social sciences. There is an overwhelming consensus that most of the kinds these sciences study and generalize about are *not* individuated by real essences and are *socially constructed* rather than natural. Some examples are as follows: "individual" and "market" in economics, "Black" and "White" in race studies, "capitalist city" in urban sociology, "woman" in feminist political science and sociology, and "emerging adult" in developmental psychology.

When saying that essentialism about these categories is mistaken, most critics mean something like the "absolute" and "categorical" essentialism that Ellis favors for physics and chemistry. However, they typically do not have in mind Ellis's physical and chemical levels. One of the alternative levels they sometimes have in mind is the genetic level of biology. Any real essence distinctions here are widely thought to fail to account for the boundaries of the mentioned social kinds. This can be called failure of *categorical real genetic essentialism about social kinds* (CRGESK). CRGESK is a nonstarter for some social categories. Nobody ever thought that a genetic distinction accounts for the category "middleman." But for other categories, such as "White male," the failure of CRGESK is more interesting. This is because the best reason given for this failure leaves open other real essentialist accounts of some social kinds, and some authors note that several debates in social sciences are clarified once we appreciate this.

The best reason to reject CRGESK concerns variation and explanation. For any social kind, there are no genetic properties that both (a) are shared by all human members of the kinds and (b) explain behaviors common to those members. The evidence for this is inductive, from genetics and population studies. It entails that for social kinds, there are no genetic properties that could meet the necessity and explanatory conditions that Ellis places on real essential properties.

This leaves open two (combinable) options for alternative sorts of *biological real essentialism* about some social kinds. First, one can expand the candidate's essential properties to biological ones other than genetic ones. For instance, there is some evidence for common and explanatory neurological properties that may individuate *economic individual* as a real kind. These properties may be intrinsic properties of human persons. But as noted above, essential properties need not be intrinsic for the nonfundamental kinds that sciences study. Races are conceived as such kinds when authors argue that extrinsic, genealogical properties of people account for the race distinctions between them.

Second, authors such as Ron Mallon have argued that the social kinds in question are HPC kinds. This allows that the properties individuating them, whether neurological, genealogical, or of other sorts, need not be possessed by all kind members. The prevalence of variation among members of each social kind would seem to require this modification of any essentialist account of them, whether biological or not. Whether authors *call* the resulting view a new form of "essentialism" is beside the point.

Social Constructivism and Biological Properties

Authors have challenged the above-described essentialist move from genetic to other sorts of biological properties. However, authors are now realizing that the resulting debates between the biologically inclined and their critics are often ill formed. This happens when the socially constructed kinds to which the critics refer and those to which the biologically inclined refer are not coextensive. Consider the following: What a biological taxonomist refers to by using race names is sometimes not what the social constructivist has in mind. Although this suggests that parties to some of these debates are talking past one another, a different interpretation is that they tacitly have a normative disagreement about which kinds social scientists *should* focus on and analyze. Uncovering these tacit normative disagreements has motivated some parties to these debates to change their argumentative strategies, focusing on crucial normative points without epistemic or metaphysical distractions. This is one place, for instance, where practical concerns about the political dangers of *applications* of biological essentialisms have a traction that they *cannot* have when it is the mere truth of those essentialisms that is in question.

But even liberalized biological essentialisms often seem to fail to account for the kinds that interest social scientists, because, as social constructivists argue, the explanatory causes of the boundaries of these kinds are *social*. This biological versus social issue is primarily empirical. For instance, it is an empirical platitude that oppressive political systems are important causes of some of the gender and race distinctions that social science recognizes.

It is crucial to note, however, that social constructivism along these lines is consistent with real, social (rather than or also biological), and HPC essentialist accounts of some social kinds. The *realism* in any such essentialism only requires that the properties that distinguish kinds are or correspond with the actual causes of kind distinctions. It does not matter whether the actions of *people* and *social groups* are among these causes. The only social constructivism that is incompatible with a liberalized real essentialism is the extreme sort, on which social kind distinctions derive *only from the mere beliefs* (not actions) of social theorizers.

The main motivation for pursuing liberalized, real essentialist accounts of social kinds, whether biological or social, is to provide a socially sensitive scientific realist's ground for well-confirmed generalizations that range over those kinds. These generalizations nearly always have exceptions, as expected on a suitably liberalized essentialism. But some of the versions of metaphysical essentialism described above allow a realist interpretation that permits exceptions. And, of course, other avenues to a realist interpretation may be open without utilizing essentialism at all. Likewise, if realism fails, this may or may not be related to essentialism.

Psychological Essentialisms

Unlike the philosophical essentialisms discussed above, psychological essentialisms concern which essentialisms (if any) people tend to believe or imply, regardless of which of these beliefs are true or justified. Psychological essentialisms still have philosophical aspects and applications though, and the social sciences, including cognitive anthropology, ethnography, and various fields in psychology, often study them.

The "classical view" of the psychology of everyday-category concepts says that we use these as though we define them by tacit appeal to sets of singly necessary and jointly sufficient conditions for instantiating the concepts. This suggests that most people are *metaphysical essentialists* of one stripe or another about everyday categories. Experiments beginning in the 1970s initiated several attacks on and alternatives to this view. For instance, researchers have suggested that instead we are (sometimes from early childhood) *presumptive scientific essentialists* about at least some categories, such as biological ones or race and ethnicity categories. That is, we assign things to categories on the presumption that there is some set of underlying, typically unobservable, features that they uniquely share and that causally explain many of the observable features characterizing the category. Authors often claim that this widespread "essentializing tendency" is innate and evolved, part of a strategy to infer generalizations from experience.

Nick Haslam and Jennifer Whelan reviewed a host of empirical studies that argue that many people are scientific (and often genetic) essentialists in this way about several social categories, including "AIDS patient," "Jews," "gay man" (and other sexual orientation categories), gender categories (especially "woman"), personality categories, race and ethnicity categories, and categories of mental disorder. Two clusters of philosophical points about these studies are noteworthy.

First, critics have noted that many of these studies are not appropriately sensitive to the sorts of distinctions between the essentialist claims highlighted above. Although some psychologists, such as Frank Keil, have tried to test between crude essentialisms and nuanced essentialist positions (such as the HPC kinds view) with respect to folk beliefs about everyday objects and biological kinds, research on social categories has not reached the same level of conceptual sophistication. There is a concern that this has compromised its conclusions.

Second, many of the studies of folk scientific essentialism about social kinds purport to show that scientific essentialist thinking about some social categories generates stereotyping and prejudice. For instance, Nick Haslam and Sheri Levy find a correlation between (a) the essentialist belief that the category "gay man" is discrete and (b) antigay attitudes toward gay men. They note that researchers often interpret this sort of result to indicate that certain aspects of essentialist thinking about the category "gay man" are sources of antigay

attitudes about gay men and that this helps account for those attitudes. However, it is difficult to find anything in such studies that justifies the inference *from* a correlation between aspects of essentialist thinking and prejudice *to* the claim that aspects of essentialist thinking cause or explain prejudice. For instance, to adapt one of Nick Haslam's and Sheri Levy's own remarks, we need finer-grained empirical work to tell whether some essentialist responses made by study participants are (however irrational) post hoc defenses of prior or otherwise caused prejudice.

Suppose that such further studies happen to justify the causal claims about essentialist thinking. A further caution is known to arise. Researchers sometimes appeal to such causal claims to support negative assessment of essentialist beliefs. Ironically, this appeal commits the same sort of naturalistic fallacy that the researchers are tacitly or explicitly criticizing. From any putative essentialist facts about social categories, no justifications of normative prejudice follow. Likewise, from any putative facts about essentialist thinking causing prejudice, no justifications of normative dismissal of essentialist belief follow. In the relevant cases, social policy would be better justified and probably practically more effective if it corrected the object of its concern, shifting from essentialist belief to dubious inferences from essentialist belief.

Matthew J. Barker

Author's Note: Thanks to Daniel Hausman and Elliott Sober for helpful comments on a draft of this entry.

See also Biology and the Social Sciences; Induction and Confirmation; Instrumentalism of Scientific Theories and Constructive Empiricism; Kinds: Natural Kinds Versus Human Kinds; Laws of Nature; Metaphysics and Science; Prejudice and Stereotyping; Race, Theories of; Realism and Anti-Realism in the Social Sciences; Reductionism in the Social Sciences; Social Constructivism

Further Readings

Barker, M., & Wilson, R. (2010). Cohesion, gene flow, and the nature of species. *Journal of Philosophy, 107,* 61–79.

Ellis, B. (2002). *The philosophy of nature: A guide to the new essentialism.* Montreal, Quebec, Canada: McGill-Queen's Press.

Haslam, N., & Levy, S. (2006). Essential beliefs about homosexuality: Structure and implications for prejudice. *Personality and Social Psychology Bulletin, 32*(4), 471–485. doi:10.1177/0146167205276516

Haslam, N., & Whelan, J. (2008). Human natures: Psychological essentialism in thinking about differences between people. *Social and Personality Psychology Compass, 2*(3), 1297–1312. doi:10.11111/j.1751–9004.2008.00112.x

Kripke, S. (1972). Naming and necessity. In G. Harman & D. Davidson (Eds.), *Semantics of natural language* (pp. 253–355). Dordrecht, Netherlands: D. Reidel.

Mallon, R. (2006). "Race": Normative, not metaphysical or semantic. *Ethics, 116*(3), 525–551.

Nussbaum, M. (1992). Human functioning and social justice: In defense of Aristotelian essentialism. *Political Theory, 20*(2), 202–246.

Putnam, H. (1975). *Philosophical papers: Vol. 2. Mind, language, and reality.* Cambridge, England: Cambridge University Press.

Robertson, T. (2008, Fall). Essential vs. accidental properties. In *The Stanford encyclopedia of philosophy.* Retrieved from http://plato.stanford.edu/entries/essential-accidental/

Sayer, A. (1997). Essentialism, social constructionism, and beyond. *Sociological Review, 45*(3), 453–487. doi:10.1111/1467–954X.00073

Smith, E., & Medin, D. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Sober, E. (1980). Evolution, population thinking, and essentialism. *Philosophy of Science, 47*(3), 350–383.

Wilson, R., Barker, M., & Brigandt, I. (2007). When traditional essentialism fails: Biological natural kinds. *Philosophical Topics, 35*(1–2), 189–215.

Ethical Impact of Genetic Research

In assessing the ethical impact of genetic research, in which philosophical and social-scientific concerns merge, it is useful to discuss three different areas addressed in this entry: (1) the ethical impact that genetic research is already having on important bioethical concepts such as *informed or valid consent*, which influence the public trust necessary for the conduct of genetic research; (2) the possible impact that genetic research will have on concepts such as free will, an issue of major importance for the social

sciences, which aim to explain social action issuing from the free will of human beings; and (3) transforming into practical issues what seem to be purely theoretical issues, such as whether it is better to have a society of people who are relatively equal to each other in their natural attributes or to have a hierarchical society whose members have widely different natural attributes.

Issues of Information

Because a volunteer cannot expect any personal benefits in volunteering for medical research, whereas a patient expects to benefit from the treatment for which she is being asked for consent, it is universally recognized that the information about the risks of being a volunteer must be fuller than the information given to patients. It must include not only the medical risks, such as suffering some bodily injury, but also risks to one's privacy and the risk of finding out facts about oneself that one would prefer not to know. In the cases with which we are concerned, a researcher is often unable to provide information about these latter kinds of risks. Genetic research, such as gene transfer research and pharmacogenomic trials, does involve risks of bodily harm, but our standard human subject safeguards apply to these.

It is becoming increasingly common, however, to create *biorepositories* of genetic information, both where the content is collected with one's consent to research and where it is obtained from existing specimens, such as from hospitals or from state newborn-screening programs. In neither case is the future use of that genomic data known. Scientific interest in using available genetic data to identify genotypic–phenotypic or environmental associations may lead researchers to simply go looking for new information. If the research identifies genetic information that meets a threshold for clinical utility, questions arise as to whether such results should be returned to the person or become the basis of recruitment into a new study. Information about a person, say, having a medical condition or being at risk for it that standard practice would not currently disclose, such as in pediatrics, where carrier and adult-onset conditions would typically be disclosed only upon reaching majority, may now be disclosed. The current efforts to craft "global" or "broad" consent forms is one way by which genetic research has already had an impact on the bioethical concept of informed consent.

Another way in which the concept of informed consent is affected by genetic research is that genetic information about an individual also provides genetic information about some of the relatives of that individual, and in small isolated societies, it may even provide genetic information about all the people in that society. These other people may not have consented to have this information about themselves provided to researchers. Thus, a researcher may gain information about people who were not even asked to participate in the research and so were not provided with the information that is required for obtaining informed consent to participate in the research project. Their privacy, as well as that of the original individual, may be violated in the use and sharing of such genetic data. Because the genetic information obtained by genetic research may have monetary value, the question has arisen whether volunteers should be provided with some compensation for their genetic data.

Issues of Determinism and Free Will

The second area where genetic information is likely to have an ethical impact is on people's attitudes. People may suffer a stigma when they are discovered to have genes associated with mental disorders or mental disabilities, even though they do not exhibit any symptoms of either. This may result from regarding these conditions as being genetically determined, so that the person is regarded as having the condition even in the absence of symptoms. Researchers claim that all genetic research is for medical purposes, such as discovering the genetic contribution to various diseases and discovering genetic variations that account for why some medications work for some people and not for others. However, it is quite clear that the genetic contribution to normal behavior will also be discovered, with potential for stigma—for example, by identifying persons prone to risky behavior. We already know that the complex herding behavior of sheep dogs is inherited, so it seems quite likely that much complex human behavior also is inherited. This may have a significant impact on people's attitudes toward free will, changing people's attitudes toward holding individuals responsible for their behavior and affecting societal attitudes toward punishment. An emphasis

on genetic associations as central to all conditions and behaviors may also cause persons to regard that genetic information as the most salient, increasing the scope of various medical conditions to permit a greater range of phenotypic expression, resulting in, for example, an enlarged autism spectrum disorder and "mild" fragile X syndrome.

Altering Human Nature

The third area in which genetic research may have a significant ethical impact is more speculative. If genetic research shows that we can manipulate the genetic makeup of an embryo so as to create a human being with new and desirable traits, it may have a significant impact on the concept of *human nature*. The concept of human nature as something given may change, so that rather than working to make the environment more hospitable to human beings, some may want to make human beings more adaptable to the existing environment. We also will have to consider as a practical matter, not just a theoretical one, whether we want a society where everyone has the kind of genetic traits that now only a favored minority have. We would not purposely create less favored individuals, as was done in Aldous Huxley's dystopic 1932 novel *Brave New World,* but not every embryo would be genetically enhanced. Thus, genetic research may have an impact on people's attitude toward equality in its most fundamental aspect, equality in desirable genetic endowment.

Bernard Gert and Arlene Davis

See also Experiments in the Social Sciences: Ethical Issues; Eugenics, Old and Neoliberal Theories of; Free Will in the Social Sciences; Genetic Indeterminism of Social Action; Transhumanism and Human Enhancement

Further Readings

Bailey, D., Skinner, D., Davis, A. M., Whitmarsh, I., & Powell, C. (2008). Ethical, legal, and social concerns about expanded newborn screening: Fragile X syndrome as a prototype for emerging issues. *Pediatrics, 121,* e693–e704.

Evans, J. P., Burke, W., & Khoury, M. (2010). The rules remain the same for genomic medicine: The case against "reverse genetic exceptionalism." *Genetics in Medicine, 12,* 342–343.

Gert, B., Berger, E., Cahill, G., Clouser, K. D., Culver, C., Moeschler, J., & Singer, G. (1996). *Morality and the new genetics*. Sudbury, MA: Jones & Bartlett.

Green, E., Guyer, M. S., & National Human Genome Research Institute. (2011). Charting a course for genomic medicine from base pairs to bedside. *Nature, 470,* 204–212.

Kaye, J., Boddington, P., de Vries, J., Hawkins, N., & Melham, K. (2010). Ethical implications of the use of whole genome methods in medical research. *European Journal of Human Genetics, 18,* 398–403.

Mitchell, R., Conley, J. M., Davis, A. M., Cadigan, R. J., Dobson, A. W., & Gladden, R. (2011). Genomics, biobanks and the trade secret model. *Science, 332,* 309–310.

Ethno-Epistemology

Western philosophy standardly defines *epistemology* as reflection upon the nature, goals, norms, concepts, and scope of evidence, justification, and knowledge per se and as evaluation of human cognitive activities from the standpoint of these norms and goals. *Ethno-epistemology* seeks a scientific understanding of epistemological reflection and evaluation. It views epistemology as a contingent and historically emergent natural phenomenon produced by a particular species of primate (*Homo sapiens*); a phenomenon that is, moreover, fully amenable to understanding from a broadly anthropological, a posteriori perspective that incorporates the relevant findings of cognitive psychology, linguistics, sociology, anthropology, evolutionary biology, and so on. The nature, aims, norms, theories, and concepts of epistemology—like those of morality, religion, science, and law—emerge from the concrete life circumstances in which epistemological inquiry is organically rooted and sustained. Epistemology, in short, is made *by* humans *for* humans. Ethno-epistemology accordingly approaches human epistemological practices without appeal to divine imperatives, transcendent norms of rationality per se, or nonnatural epistemological concepts, properties, or principles. As a scientific enterprise, ethno-epistemology extends the epistemology of the sciences (natural and social)—that is, their a posteriori evidential practices, styles of reasoning, and modes of explanation—as well as their substantive findings to the study of epistemological inquiry.

The scope of ethno-epistemology is universal in several senses. First, it examines the epistemological activities (i.e., the intuitions, judgments, norms, and goals—be they implicit or explicit) of ordinary people, of cognitive specialists (e.g., shamans, priests, jurists, linguists, cognitive psychologists, and physicists), and, most significantly, of professional philosophers, in particular professional epistemologists. This last involves subjecting to the anthropological gaze the epistemological activities of members of professional philosophical societies (e.g., the American Philosophical Association) as well as editors and referees of professional philosophy journals. Ethno-epistemology examines professional boundary work, such as how the dominant epistemological canon is constructed, how the contents of university epistemology textbooks are determined, who is deemed a "serious" epistemologist, what is deemed a "genuine" epistemological problem, and which epistemologists are mainstreamed in university epistemology courses and which are ghettoized in world philosophy, wisdom literature, or anthropology courses. It also examines professional gatekeeping activities, such as editorial decisions by professional journals, grant-awarding decisions by funding institutions (such as the National Endowment for the Humanities), admissions and degree-granting decisions by PhD-granting institutions, and academic hiring and promotion decisions.

Ethno-epistemology is universal in the second sense of viewing Western epistemology as one among many possible, contingent epistemological undertakings pursued by human beings. This approach decenters and provincializes the aims, norms, problems, intuitions, and conclusions of Western epistemology since it no longer regards these as inevitable, universal, or definitive of the epistemology standpoint per se. In so doing ethno-epistemology forsakes the double standard that exempts domestic (Western) epistemological practices from the same anthropological scrutiny and explanation that nondomestic (non-Western) epistemologies standardly receive. It likewise abjures the related double standard that privileges the epistemological activities of Western philosophers as "epistemology simpliciter" while marginalizing those of non-Western philosophers as mere "ethno-epistemology." Western epistemology is simply one ethno-epistemology among many, including pre-Han classical Chinese (e.g., Confucian), Mesoamerican (e.g., Aztec), and African (e.g., Yoruba) epistemologies. Ethno-epistemology thus resists converting being different from Western epistemology into being inferior to Western epistemology or, worse, into not counting as epistemology at all. It rejects the Hegelian evolutionary myth of the unfolding of reason in history, along with the Comtean- and the Darwinian-like social-scientific evolutionary myths of a single path of human intellectual development. There is no law of reason, thought, culture, or psychology requiring that all peoples follow the same path of epistemological development. Non-Western epistemologies do not represent earlier stages of Western epistemology.

Ethno-epistemology adopts the four methodological principles of the strong program of sociology of knowledge. They are as follows:

1. *Causality:* Ethno-epistemology examines the causal conditions governing all epistemological reflection and evaluation as well as the formation of belief and knowledge.
2. *Impartiality:* Ethno-epistemology explains both true and false belief, justified and unjustified belief, rational and irrational belief, and knowledge and ignorance in causal, a posteriori terms.
3. *Symmetry:* Ethno-epistemology invokes the same kinds of causes when explaining true and false belief, justified and unjustified belief, and so on.
4. *Reflexivity:* Principles (1) to (3) apply to ethno-epistemology itself. Ethno-epistemology is amenable to ethno-epistemological understanding.

Ethno-epistemology is conducted within science as a part of science. Its symmetry, impartiality, reflexivity, and universal scope notwithstanding, ethno-epistemology remains a definitively Western scientific project—one, it may be argued, that continues to colonize non-Western epistemologies. By forcing them upon the Procrustean bed of the combined epistemological and substantive standpoint of Western science, ethno-epistemology inevitably privileges the epistemological norms, concepts, and goals of Western science to the detriment of those of non-Western epistemologies.

James Maffie

See also Epistemology; Naturalism in Social Science; Naturalized Epistemology; Postcolonial Studies; Science and Ideology; Sociology of Knowledge and Science; Strong Program in the Sociology of Knowledge

Further Readings

Barnes, S. B., & Bloor, D. (1982). Relativism, rationality and the sociology of knowledge. In M. Hollis & S. Lukes (Eds.), *Rationality and relativism* (pp. 21–47). Cambridge: MIT Press.

Bloor, D. (1991). *Knowledge and social imagery* (2nd ed.). Chicago, IL: University of Chicago Press.

Campbell, D. T. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (pp. 413–463). La Salle, IN: Open Court.

Harding, S. (1998). *Is science multicultural? Postcolonialism, feminisms, and epistemologies*. Bloomington: Indiana University Press.

Latour, B., & Woolgar, S. (1986). *Laboratory life: The social construction of scientific facts*. Princeton, NJ: Princeton University Press.

Maffie, J. (1995). Towards an anthropology of epistemology. *Philosophical Forum, 26,* 218–241.

Maffie, J. (2002). Why care about Nezahualcoyotl? Veritism and Nahua philosophy. *Philosophy of the Social Sciences, 32*(1), 71–91.

Nichols, S., Stich, S., & Weinberg, J. M. (2003). Meta-skepticism: Meditations in ethno-epistemology. In S. Luper (Ed.), *The skeptics: Contemporary essays* (pp. 227–248). Aldershot, England: Ashgate.

Ethnography, Philosophical Aspects of

Ethnography is both a subfield within social anthropology and a methodological approach used throughout the social sciences. In the first sense, ethnographies are in-depth studies of a single culture, typically written up in the form of a monograph. As a methodology, ethnography uses diverse methods to achieve a broad and systematic understanding of a particular social group. In both senses, ethnography was developed at the beginning of the 20th century as a way to study non-Western cultures. It was later adopted by sociologists to study Western, urban communities. Today, it figures also among the methodologies of business, cultural studies, education, nursing, political science, public health, and women's studies.

This entry focuses on the methodological side of ethnography, offering an overview of the epistemological and other philosophical issues that it has raised within the philosophy of the social sciences.

Methodology

Among the methods of ethnographers, participant observation is perhaps the most iconic. The researcher lives among the subjects of her study, learns their language, and interacts with them in casual, everyday contexts. Ethnographers also use extended interviews (both structured and unstructured), as well as surveys, focus groups, discourse analysis, archival research, video and audio recording, and, occasionally, (quasi-)experimental methods. The variety of methods is required because ethnography aims at a "holistic" understanding of the social group that captures the insider's perspective. Holism is the view that the different aspects of social life cannot be isolated: The meaning of one symbol or the significance of one event depends on the meaning or significance of others.

Ethnographers draw conclusions about the particular social group they study. These are generalizations of limited scope but generalizations nonetheless. They raise at least two related philosophical issues. First, what kind of evidence supports these generalizations? Second, does ethnography suppose the existence of a distinct group of people with homogeneous beliefs, values, practices, and so on? That is, does ethnography presuppose that cultures exist? In its earliest formulation, the conceptualization of participant observation relied on a particular conception of culture. Cultures (or societies), many thought, were constituted by systems of rules or norms. An inquirer came to learn this system by "playing the game." Ethnographic conclusions were justified because the generalizations were primarily about normative systems, and these systems could be learned by living with the subjects. This epistemological and ontological stance has been questioned, and many social scientists have abandoned it.

Critical Issues

One problem noticed by social scientists was that the evidential grounding for ethnographic claims is often problematic. While the ethnographer would draw

conclusions about "the Nuer" (the ethnic group in South Sudan), his or her actual contact would often be fairly limited. In a society where activities were starkly segregated by gender, for example, a male ethnographer's evidence would be largely drawn from male informants. In addition, ethnographers often gravitated to "good informants"—loquacious and articulate individuals—without reason to believe that these individuals were representative. As long as ethnographers thought of cultures as homogeneous, these problems were no more than practical problems of empirical research. There was an independent object—the culture—of which different experiences provided glimpses. Because of the relationship between the concept of culture and the epistemic justification for ethnographic methods (particularly participant observation), critique of the idea that cultures or societies are homogeneous and distinct was an important source of concern for ethnographers.

One source of the critique comes from the recognition that conflict and change are normal conditions of human society. Non-Western cultures are not living fossils. Humans are always in contact with nearby communities, and interchange of ideas and practices is the norm. Moreover, many communities are constituted by a variety of constituencies with different perspectives on common symbols, rituals, or social events. The perspectives often align with differential power among the subjects. To represent one interpretation as "the culture" is to misrepresent at least, and at worst, it reinforces a political agenda. These arguments have had a profound influence on ethnographic practice. Ethnographers have become cautious about drawing conclusions about "the culture." And they often supplement participant observation and interviews with systematic evidence-gathering techniques that are intended to capture the breadth of different beliefs and attitudes.

A more profound critique of ethnography challenges the idea that ethnographies represent an independent object—homogeneous or heterogeneous—at all. These arguments add a critique of ethnographic writing to the considerations already discussed. Classical ethnographic writing employed the genre of realism, where the author uses rich descriptive detail to create a vivid sense of the scene. This technique, they argue, is essential to the creation of the ethnographer's authority as someone who was present and can tell the story. Ethnographic monographs, however, are typically a pastiche of people, places, and events. Indeed, ethical concerns often prompt ethnographers to disguise real people or events. In spite of its pretensions, ethnography is not a more or less accurate representation of an independent reality. Rather, it is the product of a particular interaction between the ethnographer and his or her interlocutors. On this view, expanded data-gathering techniques are beside the point. Ethnography is akin to fiction.

A variety of methodological experiments have arisen in response to these concerns. Some have maintained the commitment to realism (in the philosophical sense) and tried to justify their conclusions with more care. Others have tried to decenter the ethnographer by presenting multivocal ethnographies, where the subjects appear as cocreators, or even coauthors, of the research. For all ethnographers today, research requires careful reflection on the way in which evidence and interpretations are produced.

Mark Risjord

See also Explanation Versus Understanding; Holism, in the Social Sciences; Language and Society; Relativism and the Social Sciences: From the Sapir-Whorf Hypothesis to Peter Winch; Social Anthropology

Further Readings

Clifford, J., & Marcus, G. (Eds.). (1986). *Writing culture: The poetics and politics of ethnography.* Berkeley: University of California Press.

Evans-Pritchard, E. E. (1940). *The Nuer: A description of the modes of livelihood and political institutions of a Nilotic people.* Oxford, England: Clarendon Press.

Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In *The interpretation of cultures* (pp. 3–30). New York, NY: Basic Books.

Hammersley, M. (1992). *What's wrong with ethnography?* London, England: Routledge.

Jones, J. S., & Watt, S. (Eds.). (2010). *Ethnography in social science practice.* London, England: Taylor & Francis.

Malinowski, B. (1922). Introduction: The subject, method, and scope of this inquiry. In B. Malinowski (Ed.), *Argonauts of the Western Pacific* (pp. 1–26). New York, NY: E. P. Dutton.

Roth, P. (1989). Ethnography without tears. *Current Anthropology, 30,* 555–569.

ETHNOMETHODOLOGY

Ethnomethodology has been an important field of social science in which philosophical and methodological issues about social inquiry loomed large, being an area directly descending from a specific philosophical school. This entry presents an overview of its protagonists and their ideas and explains the distinctive kind of ethnomethodological studies.

Origins

Ethnomethodology is a dissenting strand of sociology, developed by the American sociologist Harold Garfinkel (1917–2011) during the 1950s and 1960s, which remains an active and controversial area of work in the present. Its initial connection with philosophy was through phenomenology, itself the creation of Edmund Husserl (1859–1938) in the early part of the 20th century, though the most direct link was through the work of the Austrian social theorist and philosopher Alfred Schütz (1899–1959). Schütz's project, initiated in the 1920s, was to use phenomenology's methods to clarify the philosophical foundations of the social sciences. Schütz accepted the idea of sociology as the study of social action, broadly along the lines set out by Max Weber (1864–1920), but thought that Weber had not sufficiently worked out the basic assumptions he depended on. Weber's sociology is often called a *verstehende* (or "understanding") sociology, to reflect the emphasis that Weber placed on the use of the investigator's capacity for understanding the thoughts and intentions of those being studied; but Weber had not, in Schütz's estimation, sufficiently worked out the significance of the fact that "understanding" can serve as a method for sociology only because it already serves the conduct of social life—in other words, that those involved in society are capable of understanding one another.

The Protagonists and Their Main Tenets

One of phenomenology's founding purposes was to readjust the relations between our scientific and our non- or prescientific understandings, which were often imagined to be one of outright rivalry: There is the world as we ordinarily understand it, and there is the world as it really is—in other words, as it is shown to be by science. One of Schütz's main responses to this issue was to treat everyday and scientific understanding as having different roles. Science, to put it crudely, is a theoretical enterprise, while our everyday lives involve practical understandings—a grasp of how to get things done. Because of this disparity in character, the two are basically disjunctive rather than directly competitive. Schütz saw his own contribution as that of articulating more clearly the intricacies involved in adopting the theoretical or scientific attitude in the study of practical action.

Schütz's writings had a tremendous effect on Harold Garfinkel, who sought, during the late 1940s and 1950s, to show that Schütz's conception provided a viable systematic alternative to the various theories of social action that were then available (most prominently, those of Talcott Parsons [1902–1979], the then preeminent sociological theorist). One of the crucial issues that Garfinkel thought distinguished Schütz from other sociological theorists was his treatment of the notion of *rationality*. Rationality is commonly conceived as a matter of electing the most effective means for doing things, and *if one supposes* that science is the gold standard for determining what the world—the circumstances of our doings—is really like, then effective conduct will be that which accords with scientific understandings. "Rationality" becomes a yardstick for sociologists to use in the assessment of people's doings: Can what they are doing possibly be an effective way of going about things? Are their practices rational or irrational (which is still, in other contexts, a burning topic under the title of "the rationality debate")? A further result, in application to people's doings, was to conceive the character of rational action as very much akin to following a scientific type of procedure, one that involves the systematic application of very well-defined and extremely precise rules (rational action models have been back in fashion in sociology for some time). Schütz noted, however, that the notion of rationality had a life independent of science, one applied in daily life.

Garfinkel concluded that a better understanding of how people do act in social life would be gained if one gave up invidiously matching the organization of people's actual practical doings against an idealized standard specifying how they should act to satisfy the requirements of ideal scientific rationality. This conceives the organization of people's doings in negative terms and does not really answer the question

of how people do carry out their affairs in the life of their society in ways that *they* find effective. What counts for them as rational practice? For Garfinkel, this question was not to be answered in any general and a priori way but through investigation of cases to find what standards of practical rationality were applied in different areas of social life and in the local settings in which conduct was enacted. Garfinkel thus thought of ethnomethodology as essentially a program of studies. More recently, the extent and value of Schütz's influence have been debated, and reference to Maurice Merleau-Ponty (1908–1961), an important French phenomenologist, who placed more emphasis than Schütz on the embodied nature of activity, has been more prominent in Garfinkel's thought.

Garfinkel thought that the opportunity of understanding social life as a product of practical rationalities had not really been spotted by the social sciences. Other varieties of sociology had not really taken any interest in the features identifying the affairs of daily life in society—having dinner with the family, taking part in psychotherapy, filling in social security applications, selling insurance, and so on and on—as the recognizable and mutually coordinated matters that they are, their enactment as intelligible occurrence of repeatable social doings being important to people's ability to respond to each other in standardized and coordinated ways, something Garfinkel took to be essential to the possibility of social order.

Ethnomethodological Studies

Garfinkel thus initiated ethnomethodology; its name was meant to indicate its nature: the study of those methods for practically carrying out social activities that belong to the members of society.

The first exposure of Garfinkel's thought to the sociological public in the mid-1960s had an explosive impact. It seemed to many that it was a threat to the very idea of sociology, certainly to sociology as they understood and valued it, and it was responded to, at worst, as if it were just a kind of bizarre eccentricity, more often as if it was based in fairly stupid mistakes. At the same time, there were many sociologists who saw interest and promise in Garfinkel's ideas and undertook the kinds of studies they understood him to be calling for. A large number of such studies have been made, and among the variety of these have been studies of police officers working "skid row" districts, lawyers reaching for agreement on plea bargains, teachers asking pupils questions, Buddhist philosophers in live debates, software engineers writing code and working out designs, air traffic controllers managing air space, therapists identifying patient's motives, astronomers doing observatory work, orchestral musicians in rehearsal, mathematicians working on proofs, mechanics safely doing repair work, advocates contesting the admissibility of evidence in court, and shoppers in supermarkets. These studies are all concerned with the way in which doing activities involves ordering what is done so that it exhibits the patterns of standardization and reproducibility that are characteristic features of conduct in a social order.

The studies are made on the understanding that such ordering is pervasively adaptable to particular situations—social action is viewed as a "situated" matter and so in each case must be done in, and be responsive to, local and variable circumstances, thus realizing standard patterns in quite specific forms. Just as the studies are attentive to the situated details of action's organization, so too are they typically directed toward understanding the reasoning required for the competent execution of lines of action under real-world conditions and thereby toward such general and basic topics of philosophy of the social sciences as the supposed requirements of rational reasoning, those of meaning, description, rationality, inference, deduction, rule-following, fact, evidence, measurement, and proof—these are matters with which the members of society are concerned, no less than are philosophers and theorists, but those members are engaged with them entirely in relation to the everyday demands of their practical purposes rather than for the theoretical ends that the professional thinkers seek.

For example, one study focused on the nature and use of historical evidence, though not as that is thought about among professional historians but rather as it was articulated and contested in the politically contentious setting of the U.S. Congressional hearings into the—then very controversial—Iran-Contra affair of the 1980s, where attempts to establish historical facts were very much to the fore. This example should not suggest that general matters of meaning, evidence, and so on are confined to high-profile and socially important occasions; they are pervasive of society's practical affairs and can be encountered and investigated in

all kinds of ordinary, even relatively trivial, social situations (e.g., establishing whether drivers have committed speeding offenses).

Ethnomethodology's earliest studies were of the most commonplace and everyday doings, the sort of things that "anyone" could do, such as greeting another person and starting a conversation, making a purchase at a store, or joining a queue to acquire a visa. They were undertaken to emphasize that the most unremarkable and uneventful everyday occurrences were nonetheless socially organized occurrences and no less deserving of sociologists' attention and analysis than the putatively greater and more important matters of social life. Then attention began to focus more on studying activities that typically require extensive training and specialized competences, such as working as a lawyer in a court of law, carrying out laboratory investigations in a branch of neuroscience, understanding advanced proofs in mathematics, and mastering the martial arts. This strand of work relates again to stock concerns of the social sciences, particularly issues of understanding another culture, but reshapes these in line with ethnomethodology's persistent interest in how it is that activities get done, bearing in mind that in many cases, having a proper understanding of how an activity is done is a matter of being able to do it oneself. To understand mathematics as a practical activity, for example, is very much a matter of being able to do the calculations, to solve the mathematical equations, to follow the proofs, and so on, and this might, therefore, necessitate sociological researchers in this and similar areas getting themselves up to speed, even undertaking professional qualification in the specialist activities they investigate.

This latter line of thought reflects a view of what understanding social life involves that is very different from the one otherwise predominant in sociology. There, understanding is generally assumed to come from the application of methods and procedures that govern the collection and analysis of data, ones that have been especially developed for use by professionally trained social scientists. Thus, it is sociology's distinctive epistemological, ontological, and methodological conceptions that are taken to define what counts as a proper understanding of organized social life, but it is ethnomethodology's alternative supposition that the terms for "competent understanding" are set within the socially organized affairs themselves. Thus, ethnomethodology's studies are efforts at recovering the (in a broad sense) methodological understandings that are built into the socially organized arrangements making up the affairs of daily life. There is no need for specially developed methods to effect such recoveries because it is a feature of those same affairs that they are organized so that their participants can pick up on whatever methods of practical reasoning are integral to and distinctive of those activities.

Wes Sharrock

See also Embodied Cognition; Hermeneutics, Phenomenology, and Meaning; Life-World; Philosophy of Sociology, History of; Social Construction of Reality; Weber's *Verstehende* Approach

Further Readings

Francis, D., & Hester, S. (2004). *An invitation to ethnomethodology.* Thousand Oaks, CA: Sage.

Garfinkel, H. (1967). *Studies in ethnomethodology.* Upper Saddle River, NJ: Prentice Hall.

Gellner, E. (1975). Ethnomethodology: The re-enchantment industry or the Californian way of subjectivity. *Philosophy of the Social Sciences, 5,* 431–450.

Livingston, E. (2008). *Ethnographies of reason.* Aldershot, England: Ashgate.

Lynch, M., & Bogen, D. (1996). *The spectacle of history.* Durham, NC: Duke University Press.

Lynch M., & Sharrock, W. (Eds.). (2010). *Ethnomethodology* (4 vols.). Thousand Oaks, CA: Sage.

Eugenics, Old and Neoliberal Theories of

The word *eugenics* first appears in Francis Galton's (1822–1911) work *Inquiries Into Human Faculty and Its Development*. It combines the Greek *eu,* meaning "good" or "well," with *genēs,* meaning "born." Galton (1883) defined eugenics as follows:

> We greatly want a brief word to express the science of improving stock, which is by no means confined to questions of judicious mating, but which, especially in the case of man, takes cognisance of all influences that tend in however remote a degree to give to the more suitable races or strains of blood a better

> chance of prevailing speedily over the less suitable than they otherwise would have had. The word *eugenics* would sufficiently express the idea; it is at least a neater word and a more generalised one than *viviculture,* which I once ventured to use. (p. 17, footnote)

He hoped that the new evolutionary understanding brought about by his cousin Charles Darwin would lead not only to better descriptions of the human species but also to improvements of it.

This entry opens with a concise and selective summary of the history of Galton's science of improving stock. That eugenics has a terrible history cannot be denied. What is less clear is how this history matters to current debates, in particular to those about the genetic technologies that have the power to reshape human beings. On one view, the story of eugenics serves principally as a cautionary tale. Education about eugenics tells practitioners of human genetic selection or modification what they should avoid. Another view denies that eugenics is intrinsically immoral. Once purged of its multiple moral and scientific errors, Galton's science of improving human stock can give rise to morally acceptable forms of eugenics.

The entry discusses two contemporary versions of eugenics. *Liberal eugenics* is a version of the view presented as compatible with liberal political philosophy. Eugenic ideas are present also in the radical endorsement of human enhancement called *transhumanism.*

Positive and Negative Eugenics

Galton divided eugenics into two tasks. There was *negative eugenics*. Galton lamented the tendency of the nascent social welfare system of Victorian England to hinder natural selection's elimination of poor-quality hereditary material. The feckless and unintelligent were now fed and clothed and so survived and reproduced. Negative eugenics would restore and amplify natural selection's power to purge poor-quality hereditary material. It would be combined with a program of *positive eugenics* that would encourage the well born to maximize their procreative efforts. Galton claimed to find support for this combination of negative and positive measures in the millennia of improvements of livestock effected by judicious mating.

The concept of eugenics is strongly linked in popular thinking with some of the worst crimes of the Nazis. The T4 program, which involved the murder of thousands of disabled people, was explicitly justified as eugenics. It is possible to view the Holocaust as the most brutal manifestation of Nazi negative eugenics. The Nazi determination to kill both old and young makes sense in light of their ambition to purge hereditary influences. The most widely discussed instantiation of Nazi positive eugenics was the *Lebensborn* program, which encouraged members of the German armed forces to impregnate racially approved female citizens of occupied territories.

A lamentable tunnel vision has tended to confine public awareness of eugenic crimes to their worst perpetrators. Eugenics became, in the early 20th century, a very popular idea. It found advocates and practitioners from across the political spectrum. For example, the socialist playwright George Bernard Shaw (1903/1963) insisted in his play *Man and Superman* that "the only fundamental and possible socialism is the socialisation of the selective breeding of Man; in other words of human evolution . . . We must replace the man by the superman" (pp. 723–724). In a chillingly prescient address given in 1910, Shaw mentioned a "lethal chamber" as an instrument of negative eugenics. Throughout Europe, North America, South America, Asia, and Australia Galton's science of improving human stock gave rise to injustices ranging from social exclusion to involuntary sterilization. The ideological diversity of late-19th-century and early-20th-century eugenics is well documented in the writings of the historian Daniel Kevles and the political scientist Diane Paul.

Advocacy of eugenics declined after World War II. Scholars have identified many mistakes in the eugenics movement initiated by Galton. There were some distinctively moral errors. Eugenicists mistakenly believed in the objective moral superiority of certain kinds of human being as well as of certain ways of life. There is now widespread support among serious thinkers for a moral egalitarianism that denies that some kinds of human being have a moral status superior to other kinds. In addition to this, liberal political philosophers have mounted powerful arguments for a plurality of morally legitimate conceptions of the good life. Liberals view surfers, solicitors, and soldiers as guided by different life

plans, each embodying different ideas about what matters.

Another failing, linked to the above, was the suppression of procreative liberty. As defined by the philosopher of law John Robertson, procreative liberty is the important freedom to decide whether or not to have offspring and to control the use of one's reproductive capacity. Eugenicists aspired to a population-wide coordination of reproduction. They would require those judged to have inferior hereditary material to refrain from reproducing and subject those judged to have superior hereditary material to an obligation to maximize their reproductive outputs.

These moral and political mistakes were compounded by egregious scientific errors. The eugenicists viewed the sorting of people into social classes and racial groups as allocating hereditary material according to quality, hence their plan to improve human stock by managing the reproduction of easily identified groups of human beings. There was an all-too-convenient identification of repositories of bad hereditary material with victims of popular prejudice.

The error of this picture of the human species has become fully apparent in the wake of advances in human genetics. The human genome, a rough draft of which was completed in 2000, shows a shared human genetic heritage. There are no working-class, Aryan, or Jewish genomes. Genetic analysis has revealed many variants of genes that increase or reduce susceptibility to certain diseases. These genetic variants often cluster in family groups. Some diseases are linked with certain ethnic groups—cystic fibrosis is more common in people of European ancestry, and people of African descent have a higher risk of sickle cell anemia. But no class or race bias has been found—or is ever likely to be found—in genetic variants linked with the character traits that particularly interested Galton and his followers.

Modern genetics has replaced early eugenicists' simplistic pictures of the role of hereditary influences in human development. We now recognize the traits of human beings as emerging from complex interactions of hereditary influences with environmental influences. Perhaps there are subtle, difficult-to-identify, genetic influences on fecklessness and criminality. But no genetic variant can cause feckless or criminal behavior unless combined with a fairly specific collection of environmental influences. A widely discussed study of the genetics of criminal offending illustrated the context-dependent nature of genetic influences. One version of the monoamine oxidase A (MAO A) gene seemed to make children particularly susceptible to the effects of maltreatment, raising their likelihood of subsequent criminal offending. Another version of the MAO A gene appeared to have a protective effect. Children with this genetic variant who suffered maltreatment were not more likely to criminally offend.

For the reasons summarized above, Galtonian human selective breeding is not a way to increase the level of human achievement.

Is Eugenics Still Relevant?

What is the practical significance of these many mistakes for the practice of eugenics?

In their discussion of the ethics of human genetic engineering, Allen Buchanan, Dan Brock, Norman Daniels, and Daniel Wikler present eugenics as a "cautionary tale." For them, it serves as a negative exemplar, comprising a collection of moral and scientific errors that liberal advocates of human genetic selection and modification must seek to avoid.

According to another view, eugenics remains a viable enterprise. Galton's presentation of eugenics as a science is instructive. Sciences endure through revolutions in key concepts. Although biology as practiced today is vastly different from Aristotle's biology, contemporary biologists are right to trace their discipline's history back to Aristotle. We've learned that human selective breeding is not only immoral but also ineffective as a means of improving human stock. We can ask whether there are other ways of improving human beings that target hereditary material. Indeed, Galton's definition of eugenics as the science of improving stock "by no means confined to questions of judicious mating" indicates openness to other mechanisms of human improvement. The timing of Galton's writings on eugenics, in the decades after the publication of Darwin's *On the Origin of the Species*, makes it unsurprising that he chose to focus on a possible evolutionary mechanism for human improvement. It's equally unsurprising that the decades since the description of DNA by Francis Crick and James Watson have seen a profusion of proposals to *genetically* modify or enhance humans. Human selective breeding does not work. The new genetics brings the possibility of

improving human beings without making the mistakes of Galton and his immediate followers.

The term *eugenics* is, as we have seen, strongly linked in the popular imagination with Nazi crimes. Is it needlessly provocative to retain the term for a practice for which one seeks endorsement? There are good pragmatic reasons for retaining the term. Suppose that the practices to which "eugenics" refers are not intrinsically immoral. They are, nevertheless, unavoidably dangerous. Consider another practice that we think of as intrinsically moral but unavoidably dangerous. Medical experimentation on human subjects plays an essential part in establishing the efficacy of new therapies. The history of medicine contains many cases in which the rights of human experimental subjects were ignored. Those who serve on research ethics committees are prepared for their roles by learning about these past abuses.

Acknowledging the deliberate infection of human subjects with syphilis as a medical experiment helps research ethics committees remain alert to similar abuses in the research proposals presented to them. The selection or modification of human DNA is, like medical experimentation on human subjects, unavoidably dangerous. Retaining the term *eugenics* for our current practices reminds us of past mistakes. It prevents a forgetting that the introduction of a new term might enable.

Furthermore, there is no readily available alternative term that covers the same ground as eugenics. Many philosophers use the concept of human enhancement to broach issues within the traditional purview of eugenics. But the scope of "human enhancement" differs from that of eugenics. The concept of human enhancement supposes a philosophical distinction between therapy, whose purpose is to restore or preserve levels of functioning considered normal for human beings, and enhancement, which aims to boost levels of functioning beyond human norms. Eugenics requires no philosophical distinction between therapy and enhancement. Galton's plan to improve human stock contained measures that we might think of as therapeutic—he wanted to eliminate congenital illness. But he sought also to promote hereditary factors associated with better-than-normal human functioning. Some philosophers challenge the legitimacy of the distinction between therapy and enhancement. They may find the concept of eugenics to be a more suitable label for their views.

We can identify two contemporary descendants of Galton's science of improving stock. The first, *liberal eugenics*, goes under a name that makes clear its intellectual debt to Galton. The second, *transhumanism*, though a recognizable descendant, has dispensed with the label.

Liberal Eugenics

One continuation of Galton's idea is liberal eugenics. This liberal version of eugenics defines itself in contrast to the authoritarian version envisaged by Galton and his immediate followers. Galton imagined representatives of state being informed about what kinds of human beings were best empowered to manage human reproduction accordingly. Liberal eugenics requires no assertion of the moral superiority of certain types of human being. It explicitly endorses a pluralistic view of the good life. Parents can appeal to their own conceptions of the good life for guidance on what kinds of modifications would be viewed as enhancements. They could view a change that potentially shifted their child closer to their conception of the good life as an enhancement. As we have seen, Galtonian selective breeding intrudes on procreative liberty. Liberal eugenicists would, instead, extend procreative liberty. Liberals leave untouched the established procreative freedoms, adding to these the freedom to choose some characteristics for their children. No one would be required to interfere with natural reproductive processes.

This liberal view presents the freedom to make eugenic choices as analogous to other liberties. There are no absolute liberties. The freedom of speech can be restricted if there is strong evidence of significant harm. But the onus is on those who would restrict speech to present such evidence. Liberal eugenicists would restrict procreative choices likely to result in harm either to the child or to society. But they set a high threshold for proof of harm. The harms that may result from supposedly poor procreative choices must be demonstrable rather than merely theoretical.

The liberal view focuses most directly on the procreative choices of individuals. But its advocates present it as improving human stock. Preimplantation genetic diagnosis is a currently available technology that enables the selection of human embryos to be guided by their genetic makeup. A policy of making

preimplantation genetic diagnosis available to prospective parents is compatible with procreative liberty. It enables parents to make a narrow range of choices about the genetic constitution of their future children. Suppose that the technology is mainly used to avoid passing on genetic variants linked with serious diseases. Individual eugenic choices will then have the effect of improving human stock. We may therefore achieve Galton's desired outcome without selective breeding.

Liberal eugenicists identify precedents for a freedom to use genetic technologies to choose children's characteristics. We permit parents to influence their children's characteristics by choosing environmental influences. It is permissible to provide your child with after-school tuition in mathematics. Liberal eugenicists present genetic choices as morally similar to these, a comparison that draws support from the modern understanding of human development. According to the interactionist view of development, a human being results from the complex interaction of tens of thousands of genes and uncountable environmental influences. We should ask the same kinds of moral questions of genetic enhancers that we currently put to those advocating educational or nutritional strategies of enhancement.

Opposition to Liberal Eugenics

Some thinkers oppose liberal eugenics. The German philosopher Jürgen Habermas finds liberal eugenics to be incompatible with the liberal principles from which it purports to draw support. Habermas argues that a parent who uses genetic technologies to enhance a child wrongly assumes "coauthorship" of the child's life. A parent's enhancing ambitions present to the child in a way that places them beyond challenge. Habermas opposes the comparison of genetic with environmental interventions. He argues that there is typically the option of resisting a parent's selection of environmental influences. A determined refusal to participate in tennis lessons should thwart parental dreams of sporting stardom. Parents may use bullying or bribery to overcome a child's resistance, but these measures do not eliminate the very possibility of opposition. According to Habermas, genetic enhancement works differently. A child has no say on how her early embryo got modified. When she acquires the capacity to make rational choices about her life's direction she is powerless to alter the fact that her embryo was altered. Habermas envisages that the child is forced to speculate about the origin of her desires—whether they are truly her own or are manifestations of her parents' plans for her. This asymmetry between the genetic enhancer and the genetically enhanced that Habermas claims to find undermines the egalitarian aspirations of liberal societies. In effect, today's genetic enhancers assume an illegitimate power over future generations.

A further challenge to liberal eugenics comes from the political philosopher Michael Sandel. Sandel's critique focuses on the liberal comparison of genetic with environmental enhancements. He proposes that the correct comparison is not with normal, healthy parenting but with managed, high-pressure child rearing, properly labeled "hyperparenting." Hyperparents pay little or no heed to the nascent interests of their children. They obsessively manage children's environments to turn them into academic or sporting stars. According to Sandel, hyperparenting exemplifies an unhealthy attitude to children, and so does genetic enhancement. Healthy parenting involves accepting that there is much about our children that is not up to us. This includes the formation of our children's genomes. Hyperparents view their offspring as objects of transformation. Genetic enhancement both facilitates and exaggerates this tendency.

These challenges have elicited responses. Habermas may overlook opportunities to resist the plans of parents who seek to genetically enhance their children. While you cannot make it the case that your embryonic DNA had never been altered, you can prevent the alterations from having the effects your enhancer was hoping for. According to the interactionist view of development, significant traits emerge not from the action of genes alone but from the interaction of genes with the environment. A child can successfully resist her parent's plans by refusing to place the modified gene or genes in the environment that they require to produce their intended effects. Suppose that you learn that your genome was altered with the aim of turning you into a champion rugby player. Such a plan requires a high degree of cooperation from you. If you refuse to ever set foot onto a rugby field, you are unlikely to excel at the sport. One response to Sandel accuses him of overstating the power of genetic enhancers. A realistic depiction of genetic enhancement supposes

control only of a small subset of the total collection of genetic influences. Parents who introduce into their child's genome a genetic variant linked with sporting success can claim to accept much about their children. Perhaps they will compensate for their limited intervention in their child's genetic makeup by adopting a less intrusive approach to her upbringing.

Transhumanism as a Form of Eugenics

Transhumanism is a movement whose goal is to eliminate aging and to greatly enhance human intellectual, physical, and psychological capacities. Transhumanists are vigorous advocates of human enhancement. Their vision of our future combines millennial life spans with dramatic multiplications of our cognitive powers. Transhumanists reject the suggestion that concern about the remaining humans might place limits on enhancement.

Is this eugenics? The term *transhumanism* was coined by the British biologist Julian Huxley in an article published in 1957 that urged that human existence undergo transcendence, shifting from "a wretched makeshift, rooted in ignorance," toward "a state of existence based on the illumination of knowledge and comprehension" (p. 16). In the years before World War II, Huxley had been a strong advocate of Galtonian selective breeding. After the war, his enthusiasm for the concept of eugenics waned. Modern transhumanists replace Galton's commitment to selective breeding with an interest in a variety of transformative technologies or therapies. Many of the technologies that interest them bypass human hereditary material altogether. For example, the futurist Ray Kurzweil advocates enhancement by grafting a variety of cybernetic implants to our bodies and brains. Transhumanists advocate no restrictions of procreative liberty. They believe extended life spans and enhanced cognitive powers to be so attractive that no violations of liberty will be required. Transhumanists do not view the members of certain social classes or races as superior to others. Indeed, they tend more toward a somewhat dismissive view of distinctively human characteristics. For transhumanists, our humanity is raw material fit for refashioning into radically improved forms.

The transhumanists want changes that are much more dramatic than those that could be achieved either by selective breeding or by the cautious modification of human DNA envisaged by liberal eugenicists. Some transhumanists envisage us changing to such an extent that we will no longer be human. We will be *posthuman*. This call for individuals to enhance to the point at which they are, or may be, no longer human has forced their opponents to clarify exactly what of value might be at stake in the loss of our humanity. There's an irony in eugenics, a movement whose original purpose was to improve human stock, giving rise to transhumanism, a movement that may lead to the abandonment of our humanity.

Nicholas Agar

See also Biology and Social Sciences; Ethical Impact of Genetic Research; Genetic Indeterminism of Social Action; Sociobiology; Transhumanism and Human Enhancement

Further Readings

Agar, N. (2004). *Liberal eugenics: In defence of human enhancement.* Oxford, England: Blackwell.

Agar, N. (2010). *Humanity's end: Why we should reject radical enhancement.* Cambridge: MIT Press.

Buchanan, A., Brock, D., Daniels, N., & Wikler, D. (2000). *From chance to choice: Genetics and justice.* Cambridge, England: Cambridge University Press.

Galton, F. (1883). *Inquiries into human faculty and its development.* London, England: Macmillan. Retrieved from http://galton.org/books/human-faculty/text/html/index.html

Habermas, J. (2003). *The future of human nature.* Cambridge, England: Polity Press.

Huxley, J. (1957). Transhumanism. In *New bottles for new wine* (pp. 13–17). London, England: Chatto & Windus.

Kevles, D. (1998). *In the name of eugenics: Genetics and the uses of human heredity.* Cambridge, MA: Harvard University Press.

Kitcher, P. (1996). *The lives to come: The genetic revolution and human possibilities.* Middlesex, England: Penguin.

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology.* London, England: Penguin.

Paul, D. (1995). *Controlling human heredity: 1865 to the present.* Atlantic Highlands, NJ: Humanity Books.

Robertson, J. (1994). *Children of choice: Freedom and the new reproductive technologies.* Princeton, NJ: Princeton University Press.

Sandel, M. (2009). *The case against perfection: Ethics in the age of genetic engineering.* Cambridge, MA: Harvard University Press.

Savulescu, J. (2001). Procreative beneficence: Why we should select the best children. *Bioethics, 15,* 413–426.

Shaw, G. B. (1963). Man and superman. In *Complete plays with prefaces* (Vol. 3). New York, NY: Dodd, Mead. (Original work published 1903)

The transhumanist FAQ (collated by N. Bostrom). (n.d.). Retrieved from http://humanityplus.org/philosophy/transhumanist-faq

Wasserman, D. (2003). My fair baby: What's wrong with parents genetically enhancing their children? In V. Gehring (Ed.), *Genetic prospects: Essays on biotechnology, ethics, and public policy* (pp. 99–110). Lanham, MD: Rowman & Littlefield.

Events

This entry introduces the philosophical discussion on events in metaphysics, and goes on to point out the status of events and the role they can play in social ontology and social explanation. The importance of events in the philosophy of the social sciences lies, primarily, in that once their status is assured and their elements shown to be irreducibly social items (collective entities, social properties, etc.), this can be seen to bear directly on the holism–individualism debate in the philosophy of the social sciences.

Introduction

Events, as a special ontological category, have been the subject of systematic philosophical discussion only relatively recently, while they have been rather ignored, as such, by the philosophy of the social sciences. Yet we naturally talk, assess, or debate about social, political, and historical events in our everyday talk; social-scientific discourse, too, refers to, analyzes, and tries to provide explanations for social events, or classifies them into kinds and appraises their importance, especially in studying history, politics, or international relations. Virtually all social-scientific subdisciplines employ terminology that refers to events in one way or another, yet hardly any sustained theory about them has been offered in the philosophy of the social sciences. Unlike the latter, by contrast, even if only recently, philosophy has turned its attention to events as denizens of ontological domains, that is, as a genuine topic in metaphysics.

The discussion that follows covers, first, some of the most important issues focused on by this recent philosophical concern with events in metaphysics and then moves to the special case of social events and suggests a place for them in the philosophy of the social sciences. An obvious point where the metaphysics of events meets the social world is human action; philosophical theories of action and agent-causation have already been developed, but for the most part, this analysis has proceeded independently of the philosophy of the social sciences as such. These issues are treated separately in this encyclopedia primarily in the entries on the philosophy of action, agency, collective agency, and on causes versus reasons in the explanation of actions.

Events as an Ontological Category

The debate about events as an ontological category is spurred by the need for a special category that would help solve problems in a variety of philosophical areas: time and change; causation (What are the relata of a causal relation—events, facts, things, or something else?); action theory (Are actions a species of events? What is the relationship between actions and bodily movements? Can there be agent-causation as distinct from standard event-causation?); the mind–body problem (What is the relationship between mental events and physical events in the brain?); the structure of explanation; the logical form of action sentences involving adverbial modification; truth-making theory; and so on. In addition, science, and in particular physics (especially quantum mechanics) and biology, imposes certain requirements on what can count as an event, and therefore, scientific facts have to be taken into account rather than being disallowed by metaphysical theories or armchair theorizing. Thus, science may be allowed to have a legitimate influence on the philosophical analysis of events. The same holds for the social sciences. To the extent that they supply us with a number of (kinds of) events or a typology of social and historical events, the social sciences cannot be disregarded if a special ontology of social events is to be forthcoming.

The usual strategy in philosophical discussions about events has been first to take a position in favor of or against countenancing talk of events and then, if they are accepted, to give a theory or account of what sort they are.

Events, Facts, and States of Affairs

A natural starting point for deciding whether events constitute a separate ontological category is the

debate on whether events should be distinguished from kindred categories such as "facts" and "states of affairs" ("states" and "processes," though certainly not to be collapsed into the others, can for the present purposes be treated as similar sorts or as ongoing events or unchanging processes, in which case the notions of "event" and "change" are not necessarily or not always linked with each other, as our intuition would lead us to believe).

Earthquakes, elections, engagements, emissions, eruptions, and escapades are normally considered events in everyday talk. The first point that comes to mind, pre-analytically, is that events are happenings or occurrences and, hence, are things that can be said to "happen," "take place," or "occur" (and for some, though disputed, things that may "recur"), not something that "exists." For this reason, it is widely concurred that it is best to say that events *are* (rather than exist), if one decided to admit them as a separate ontological category. The second point that comes to mind is noticing how language effortlessly allows plural expressions of events—e.g., an earthquake as well as earthquakes, an evacuation as well as several evacuations—hence for those in favor of events as a bona fide category, such an ability to quantify shows that events are particulars (of a sort to be explained shortly) and therefore metaphysical items separate from facts or states of affairs, about which it is obviously quite unnatural to talk in plural constructions. "The earthquake that shook Shaanxi in 1506 was the deadliest ever" and "Yesterday's national elections in Erewhon had a record high in turnout" are different linguistic constructions compared to those employed to talk about facts, such as "the fact that the earthquake that shook Shaanxi in 1506 was the deadliest of all" or the embedded that-clause in the following sentence: "that yesterday's national elections in Erewhon had a huge turnout was a welcome development in that country's recent political life." Similarly, the linguistic differences between the phrases "Vesuvius's eruption" and "Vesuvius's erupting" are considered indications signaling a significant ontological difference between what is named by them: What is named in the former is a particular thing of a certain sort, an event, while in the latter something else.

While events are said to "happen" or "occur," states of affairs are said to "obtain" or "not obtain" and if the former, they are considered facts or "what is the case." In the sentence "The exposed radioactive materials were responsible for *Beta-particle emission* within the spacecraft," the italicized noun phrase names what happened—an event—while in "The postmortem scientific investigation showed that *the fact that beta-particles were emitted* was caused by exposed radioactive materials in the spacecraft," the italicized phrase refers to *that such and such was the case,* that is, to a state of affairs that obtained. Since the latter is something that is true or that is the case, it has been natural to think that states of affairs (and indeed those that obtain and are thus facts) must be expressed by true sentences or that they are identical to true propositions. Indeed, some philosophers have considered facts (or states of affairs that obtain) to be isomorphic to propositions or as what make propositions true (cf. the recent topic of truthmakers).

So it is not surprising that in recent discussions in analytic philosophy, a semantic analysis of event talk has been seen as a suitable way to make successful inroads into the metaphysics of events and as a useful way to distinguish between events and facts or states of affairs. In the examples above, we can discern what is called nominalization, that is, the construction of noun-phrases such as "yesterday's elections" or "Vesuvius's erupting"; the former have been called *perfect nominals* while the latter (containing gerunds) *imperfect nominals*. The linguistic differences between the two have been seen as reflecting a distinction between event-language versus fact-language; on this approach, syntactic and semantic differences indicate corresponding differences in the metaphysics of events and states of affairs/facts, respectively: only perfect nominals can replace a name or description of the subject of an event by a definite or indefinite article ("Vesuvius's destruction of Herculaneum" can be rephrased as "The destruction of Herculaneum" whereas "Vesuvius's destroying of Herculaneum" cannot; "radioactive materials' emission of β-particles" can be transformed into "the emission of β-particles" but not the noun-phrase "radioactive materials' emitting of β-particles"); only perfect nominals permit plural forms, as we have seen already, adjectives can be used attributively for perfect nominals whereas imperfect nominals are modified by adverbs (add, e.g., the words *slow* and *slowly* in our pairs just mentioned); and so on. Thus, perfect nominals name events while imperfect nominals name facts.

Syntax or the semantics of naming is not of course a convincing approach for those who dispute the

existence of events. One thing the aforementioned linguistic route to events shows is that, unlike facts or states of affairs that can be said (on some views) to be universals—existing necessarily, never existing necessarily (as with contradictions), or always existing but not necessarily (i.e., contingently), though they may obtain or fail to obtain—events are locatable and datable particulars: Events occur at specified places and times and, hence, are unrepeatable occurrences. Whether events are abstract or, in contrast, concrete particulars on a par with enduring objects, or so-called continuants, is another, disputed issue (leading to the debate whether more than one event can collocate at the same time, something important for action theory and social ontology—see below). At least the consensus would have it that, as long as events are not to be conflated with concrete particulars, like chairs and ballot slips, events are not like facts or states of affairs when it comes to their location in space and time: facts can be seen to be atemporal and nonspatial. But even though events are within space-time, as it were, they are not so in the same way that ordinary material objects are. The latter endure through time, occupying (perhaps different) places at different times while remaining present in all of them (they continue to exist, hence the term *continuant*), whereas events are "spread out" in time, that is, they have different temporal parts without wholly occurring in each of these time-slices. So, on the standard view (by all means not endorsed by everyone) events are said to "perdure" unlike material objects, which "endure" through time. However, if we take events to be changes in their subjects, we cannot also say that events move, as ordinary material objects do.

Events as Particulars

Two principal ways of construing events as nonrepeatable particulars have been offered: either as *structured* or as *unstructured* entities. These two views correspond to two of the most influential theories of events, Jaegwon Kim's and Donald Davidson's, respectively.

According to Kim's *property-exemplification* conception of events, an event is the having of a property by a specified object or objects at a specified time: A certain event, *e*, happens or takes place just in case a specified thing (the subject of the event, *s*) exemplifies or has some property (the constituent property of the event, *p*) at a certain point in time (the time of the event, *t*). Hence, an event, *e*, is a structured entity represented by the ordered triple $\langle s, p, t \rangle$: The earthquake that shook Shaanxi in 1506 constitutes an event since a certain particular, Shanxi province, exemplified a property, being shaken by an earthquake, on a certain day of that year. This view sees events as basically timed changes in a certain subject: John's having voted yesterday counts as a certain change in John, the subject, i.e., his exemplifying a certain property, voting. Similarly in the case of relations, e.g., "The Labour MP was reelected by the voters of the Brompton constituency last Thursday," we have two subject-items and the relation holding between them.

When it comes to providing identity conditions for events, it follows that on this theory, two events, e_1 and e_2 are identical just in case they share the ordered triple:

$$\langle s_1, p_1, t_1 \rangle = \langle s_2, p_2, t_2 \rangle \textit{ iff } s_1 = s_2, p_1 = p_2 \text{ and } t_1 = t_2.$$

For example, the Visigoths' sack of Rome in 410 CE is the same event as Alaric's army raiding Rome in 410 CE, to the extent that the constituent subjects, properties and times respectively, are identical. When properties diverge or when a property term is replaced by a different one (sacking, conquering, or liberating, for example), or when we have collective entities such as corporations or armies rather than single subjects such as a CEO or a general in the constituent subject position, then on this view of events, identity is affected—something evidently quite important for the social sciences.

On Davidson's theory, events are unstructured particulars any one of which can be described in a variety of ways, retaining its identity as a single event, even though different subjects and different properties are being mentioned in each such distinct description. Davidson's theory has been influential primarily because of its rationale, that is, what it was proposed for. Davidson wished to solve two problems by postulating events as quantifiable particulars: one, to find out what the relata of causal relations are, especially in action theory, and, second, to solve the problem of adverbial modification without going outside first-order predicate logic. "Sidney announced his resignation by reading it,

slowly yet irately, at the Quaker's annual meeting in London last Tuesday" names an event-action that is characterized or modified by a number of adverbs, of manner, place and time. If we start dropping all these adverbs one by one, does that sentence entail the ones we produce in that way, i.e., sentences like "Sidney announced his resignation by reading it slowly" or "Sidney announced his resignation by reading it in London" and so on? To do so, we must provide an account of the logical form of such action sentences as our original one, and this requires us to see them as covert existentially quantified sentences asserting the existence of an event, Sidney's announcement of his resignation, the adverbs being the adjectives of that event (that it was slow, that it involved reading a text, in a certain room, etc.). In this way, canonical notation already employed in first-order predicate logic is retained, together with the advantages of entailment preserving truth:

$$\exists e\,[(\text{Resignation announced (Sidney, } e)) \wedge (\text{Reading } (e)) \wedge (\text{Slow } (e)) \wedge (\text{Irated } (e)) \wedge \text{At } (\text{Quaker's annual meeting } (e)) \wedge \text{In}(\text{London } (e)) \wedge \text{On}(\text{last Tuesday } (e))]$$

This is our original full sentence put in logical form as a conjunction, and can be seen to entail any one of the other ones in which some or all of the adverbial modifiers are missing, since the truth of a conjunct is entailed by the truth of the conjunction containing it. So if the original full sentence is true, so is any one of its conjuncts, such as, e.g., "Sidney announced his resignation" where no other item of information is retained. Then to this sentence a number of adjectives (ex-adverbs) can feature as modifiers of the event named by it. Obviously such a proposal is significant when we come to construct explanations of social events where such information plays an important role in understanding what happened and therefore cannot be eliminated. But we should be careful here not to conflate two distinct issues; it is one thing to say that in such cases, the entailment holds, producing true sentences, and quite another to claim that we can fruitfully employ it in explaining. *Entailed* by is not (or not always) the same as *explained* by. The proposal we are discussing has to do with ontology (and truth-preserving), not social explanation, where our purposes diverge. However, it may be claimed that the proposal that uncovers a tacit quantification over entities such as events may provide fruitful notions for social ontology. One of the significant things about Davidson's proposal was that it turned the tables on the ontological priority of ordinary things, making events more significant.

Critical Issues

What are the relative merits of these two proposals —events as structured particulars versus events as unstructured ones? Problems turn first on the identity conditions for events (how to individuate events) proposed by each. Naturally, this is crucial, as we have already glimpsed, especially for the philosophy of action and the philosophy of the social sciences. For it matters greatly whether an action carried out by means of a certain bodily movement counts as one and the same event when it is described in different ways, as a socially defined or socially meaningful action in a social context or as a bodily movement; for example, is an action described as paying bus fare at the student rate the same event as someone's hand dropping coins into the fare box, or round metal things into a box? Furthermore, even if we disregard the level of bodily movement or the physical description of social objects like coins (invested with social meaning), and remain at the social level only, we can still ask whether events are the same when they are variously described, as we have seen above with the example of the sacking of Rome. When we describe an event with reference to different social entities or social properties than those figuring in the original description of that event, does this dramatically alter it so that the descriptions are no longer of the same event? Here the issue of extensionality allowing substitution of synonyms salva veritate is quite important. It also ties up with Russell's Theory of Definite Descriptions.

Furthermore, with respect to intentional action, Elizabeth Anscombe has famously suggested that when one acts with intention to do something, the action is intentional *under a description* that refers to the actor's reasons or intention; however, the same action can be seen as a physical event *under a different description*. In the latter case, it would be useless in any social explanation, yet with respect to the ontology of events (action-events here) we are concerned with, it matters equally greatly whether, explaining apart, they are one and the same event

from an ontological point of view (and it also matters equally insofar as moral responsibility is involved). Wittgenstein's followers, like Peter Winch, have made similar claims regarding the understanding of meaningful behavior within rule-following contexts (signing one's name on a piece of napkin, on a bank check or on a international peace treaty official document may require the same kind of bodily movement in all cases, though the actions [and hence the events] are distinct given the correspondingly distinct "rules of the game" within which each action is embedded so as to count as, e.g., signing a check and not as some idle pastime in calligraphy). Also, the issue of event-collocation arises here, for we have to decide whether more than one event is present when, for example, somebody makes a money transfer thereby also performing a bankruptcy.

The primary criticism of Kim's theory has been that it leads to excessive and unpalatable fine-grained distinctions. Once a slight change is registered in how any of its structural constituents is specified, we get a different event, thus multiplying events needlessly: John's pushing the ballot slip down the slot inside the voting booth is rendered quite a distinct event from John's exercising his voting rights at that time and place, once the constituent property is altered. Obviously, this is an important consequence, especially for social explanation. And it is more pronounced once we acknowledge that, as in the examples above, on a property-exemplification theory such as this, a different property makes a huge difference and equally so in the case where the subject of the events is a collective agent, such as a government, or a social-role occupier, such as a justice of the peace, or where the event-action is a joint or team action (given a team's changing membership while retaining its identity over time, a recurring action by such a team counts as the same kind only if we disregard individual membership). One way, pertinent to social explanation, to deal with these cases is to say that the different properties are instances or species of a generic property, but this leaves us with the next problem one level up, providing identity conditions for such property-hierarchies, genus and species, in a non-circular fashion. Other criticisms of Kim's view center around metaphysical issues (whether properties can be static, etc.). One way Kim defends the criticism is by drawing our attention to the distinction between a property being the constituent property of an event (one of its inner structural elements) had by the subject of the event and being a property of the whole event as such: predicate modifiers such as in our example above, by reading, slowly, irately, must be seen not as modifiers of the constituent property, resignation-announcing, exemplified by Sidney, but as characterizing the event as such, i.e., as slow, etc., hence blocking the unnecessary proliferation of events. This of course does not work with other adverbs that are not modifiers of manner or means, but adverbs in attributive positions denoting predicates stemming from legal or social norms and regulations, as for instance in "He passed a resolution illegally or against the organization's regulations" from which it follows that he did not in fact succeed at all in passing such a resolution in the end, and therefore there is no event that counts as "passing a resolution" to which the adjective *illegal* applies, for there is no such event at all. Another way of bypassing the criticism of unnecessary proliferation of events is by acknowledging a special kind of event as a "generic event" in each case (e.g., kissings, marriages, international treaties) and treating all the rest bearing adverbial modifications as subspecies.

Finally, another criticism against property-exemplification theories is that we may countenance (and some have) events that are subjectless, that is, events without any participants: Various examples of physical phenomena have been offered (and responses, too). But for our purposes, it may be interesting to note that this possibility of subjectless events, or of events lacking clearly identifiable participants may be of value to the philosophy of the social sciences. Such a possibility may indeed corroborate collective phenomena like group decisions or judgment aggregation processes, whereby no identifiable, enduring entity can be picked out as playing the role of the subject exemplifying a property (that of decision making). In addition, allowing subjectless events may corroborate structure theories as against agency theories in sociology, as discussed below.

In Davidson's case, besides criticisms regarding his specific proposal regarding the preferred logical form for taking care of adverbial modifications, which we have seen above, the most potent criticism centers around the other reason for his proposed analysis: the identity conditions of events. Two events are numerically identical, on his view, just in case they have the exact same causes and the same effects. But it is blatantly obvious that this is a

criterion inviting circularity, for unless we have such a criterion to determine sameness of events, we cannot call upon the sameness of causes and/or effects, themselves being events. Another worry is that there may well be events that lack either causes and/or effects, in which case the criterion breaks down.

In addition to criticisms centered on the proposed identity conditions in each approach, there are other concerns voiced. In Kim's case, an event being an exemplification of a property by a subject at a time involves a relation of instantiation of that property by the subject-item that has it; but this brings in the ancient-old problem of infinite regress in trying to ground a "link" such as instantiation or exemplification between an item *a* and the property *G* on yet another "linking relation," thereby bringing in continuously ever more such "links" ad infinitum.

Social Events?

Event-like items have been acknowledged in social theory since Durkheim's category of "social facts." Also, the philosophy of history has shown particular sensitivity to such entities (see, e.g., the foremost issue: Are historical events unique?). An earlier first attempt to provide a philosophical analysis of events suitable for social ontology has been offered by Kaldis (1993, Pt. III).

However, recent developments in social ontology involving theories of collective intentionality, joint action, or plural subjecthood, as well as the metaphysics of events adumbrated above, call for a renewed interest in events as a separate category over and above individual agents, collective entities, or social persons. Event ontologies where events are seen as the primary or most basic or irreducible ontological category vis-à-vis ordinary persisting things, like material objects, have been the focus of certain positions in metaphysics as well as of studies relating them to modern physics (like quantum mechanics and field theories), whereas there seems to be no comparable interest in the social sciences. Even reductionist approaches and those employing the concept of supervenience, which see events as ontologically derivative and ordinary things such as substances as primary, still allow talk of events (unless we opt for outright eliminativism), whereas yet other views allow for pluralism of ontological categories. A similar strategy may be followed in social ontology. Equally, while motion in physical space as a kind of change has been the focus of attention in discussions of events, no comparable interest has been shown in the social case, where events can be seen to be, ontologically, changes of sorts, irreducible to their individual participants. Furthermore, if change is, at least according to some views, what events should be characterized as (principally, as a change in the properties or relations of the subjects of events), then what counts as a bona fide "relational change" is a topic of particular importance to the study of social events that involve primarily such purely relational changes whereby no intrinsic change in the subjects themselves is involved.

Another issue that can throw light on social ontology and social explanation has to do with fact-causation versus event-causation. Those who reject the former by reducing it to absurdity (by so-called slingshot arguments) start by claiming that any causal relation worthy of its name must be (unlike causal explanation) purely extensional, whereby such a relation holds between entities irrespective of how they can be conceived or described by us. The distinction between explanatory relations holding in intensional contexts while causal relations remaining extensional has been usually evoked as a crucial difference, yet as mentioned above, it is a usual mistake in the philosophy of the social sciences to conflate the two or slide from the intensionality of the former to that of the latter.

In addition to the issue of intentional action under a description, which is central to social explanation and social ontology, the debate about the nature of events concerns the social sciences in many other ways as well. First, there are the two clashing answers to the issue of whether actions are events—one denying that an agent's causing of an event is itself an event, that is, an event locatable in space and time on pain of absurdity, and by contrast, the opposite view construing actions as events identifiable (on some readings) with bodily movements. The former view may be seen as placing too much emphasis on agency, construing actions as performances but events as happenings. The latter view remains devoted to what the action as such involves and not how it may be described or not as being a cause of whatever other event it brings about and that is thereby classified by its effect, as, e.g., a bank transaction. Given certain demands of explanatory contexts, the second view may go on to add the effects of

such an action-event and hence choose how to describe it. It seems that this distinction in philosophy corresponds, *ceteris paribus,* to something analogous in social science, the agency/structure dichotomy: If it matters how we describe a social event and also that it is primarily something that an agent performs, then the philosophical position that denies that actions are events is akin to agency theory in sociology, while if an action's consequences are ignored and all that matters is the event that an action is, then it is akin to structure theory in sociology.

Finally, a number of issues in metaphysics having to do with whether there can be instantaneous events, non-events (negative events: John did not kill the president), unique events, simple and complex events or modes of compositionality for events having other events as proper parts, can also be seen as important issues that are also applicable to social and historical events and as featuring both in social ontology and in social explanation.

Conclusion

Many worry whether a systematic theory of events proper can ever be constructed, arguing instead that events can be accepted as items in our ontological pictures of the world, without thereby providing any real theories for them independent of what they can be used for or in relation to other, more basic items included in those pictures. In addition many have felt more comfortable with a pluralist view of quite different event-kinds, according to which we do not need to stipulate universal ontological principles of eventhood, as it were, across the board, that is, provide a definition of *event* as such.

Byron Kaldis

See also Action, Philosophical Theories of; Causes Versus Reasons in Action Explanation; Collective Agents; Holism, in the Social Sciences; Individualism, Methodological; Mind–Body Relation; Social Facts; Social Ontology, Recent Theories of; Supervenience

Further Readings

Bennett, J. F. (1988). *Events and their names.* Oxford, England: Oxford University Press.

Bennett, J. F. (1991). Précis of events and their names. *Philosophy and Phenomenological Research, 51,* 625–628.

Brand, M. (1997). Identity conditions for events. *American Philosophical Quarterly, 14,* 329–337.

Casati, R., & Varzi, A. C. (Eds.). (1996). *Events.* Aldershot, England: Dartmouth.

Chisholm, R. (1970). Events and propositions. *Noûs, 4,* 15–24.

Cleland, C. (1991). On the individuation of events. *Synthese, 86,* 229–254.

Danto, A. (1985). *Narration and knowledge.* New York, NY: Columbia University Press.

Davidson, D. (1980). *Essays on actions and events.* New York, NY: Oxford University Press.

Davidson, D. (1985). Reply to Quine on events. In E. Lepore and B. McLaughlin (Eds.), *Actions and events: Perspectives on the philosophy of Donald Davidson* (pp. 172–176). Oxford, England: Basil Blackwell.

Dodd, J. (2009). Events, facts and states of affairs. In R. Le Poidevin et al. (Eds.), *The Routledge Companion to* Metaphysics (pp. 322–334). London, England: Routledge.

Higginbotham, J., Pianesi, F., & Varzi, A. C. (Eds.). (2000). *Speaking of events.* Oxford, England: Oxford University Press.

Horgan, T. (1978). The case against events. *Philosophical Review, 87,* 28–47.

Kaldis, B. (1993). *Holism, language and persons: An essay on the ontology of the social world.* Aldershot, England: Ashgate.

Kim, J. (1993). *Supervenience and mind: Selected philosophical essays.* New York, NY: Cambridge University Press.

Lombard, L. (1986). *Events: A metaphysical study,* London, England: Routledge and Kegan Paul.

Lombard, L. (1998). Ontologies of events. In C. Macdonald & S. Laurence (Eds.), *Contemporary readings in the foundations of metaphysics.* Oxford, England: Blackwell.

Lowe, E. J. (2002). *A survey of metaphysics.* Oxford, England: Oxford University Press.

Nowell-Smith, P. H (1956). Are historical events unique? *Proceedings of the Aristotelian Society, 57,* 107–160.

Pianezi, F., & Varzi, A. C. 1(996). Events, topology, and temporal relations. *The Monist, 78,* 89 -116.

Simons, P. (2003). Events. In M. Loux & D. W. Zimmerman (Eds.), *The Oxford handbook of* metaphysics (pp. 357–385). Oxford, England: Oxford University Press.

Taylor, B. (1985). *Modes of occurrence: Verbs, adverbs and events,* Oxford, England: Blackwell.

Thomson, J. J. (1977). *Acts and other events.* Ithaca, NY: Cornell University Press.

Quine, W.V.O. (1985). Events and Reification. In E. Lepore & B. McLaughlin (Eds.), *Actions and events: Perspectives on the philosophy of Donald Davidson* (pp. 162–171). Oxford, England: Basil Blackwell.

Vendler, Z. (1967). *Linguistics in philosophy.* Ithaca, NY: Cornell University Press.

Evidence-Based Policy

This entry introduces the idea of rigorous scientific evidence needed for devising and evaluating public policies, explains the main techniques, and reviews a number of problems that need to be overcome. Rigorous social-scientific methods, as well as philosophical issues on causation, probability, and the like, thus enter (and should concretely be taken into account in) the area of policy making. The entry also explains the relation of evidence-based policy with what has been called the "audit society."

Predictions about the *success of policy interventions* should not be dictated by fashion, ideology, political pressure, lobbying, or habit. They should be based on *rigorous scientific evidence,* if possible on results from randomized controlled trials (RCTs), and certainly not on anecdote, casual observation, theory without experiment, or expert opinion. That's what counts as *evidence-based policy* (EBP), which is now widely mandated at local, regional, national, and international levels.

The central drivers for EBP have been the United Kingdom (UK) and the United States, though it is spreading in Europe and to international organizations like the World Bank. The Blair government committed the UK to EBP in its Modernising Government White Paper of 1999. In the United States, it became entrenched through the 1996 Congressional demand for "comprehensive evaluation of effectiveness" of federally funded state and local anticrime programs and then by the No Child Left Behind Act of 2001, which allowed federal funds only for educational activities backed by "scientifically based research."

The movement borrowed heavily from the *evidence-based medicine movement* of a few years earlier, both in rhetoric and in methodology, especially in the emphasis on RCTs. In 1993, the international nonprofit Cochrane Collaboration was founded to provide systematic reviews of the best available research evidence on health care interventions; in 1999, the Campbell Collaboration was founded to do the same for education, crime and justice, and social welfare.

EBP aims to find "What Works," as in the titles of the U.S. Department of Education's What Works Clearing House and the National Institute of Justice's report to Congress, *Preventing Crime: What Works, What Doesn't, What's Promising.* There are by now large numbers of guidelines available detailing what counts as rigorous evidence and agencies, like the Campbell Collaboration, using these guidelines to determine which policies are supported by rigorous evidence.

The call for EBP marched hand in hand with the rise of the so-called audit society, with its demand for accountability and control. Interventions are to be restricted to those backed by rigorous evidence, and rigorous methods of monitoring are to be employed to ensure that policies work once in place. Audits, though, require precise, measurable outcomes. So, too, do rigorous scientific studies. But policy goals seldom have the kind of precision required. We want school leavers to be competent enough at mathematics to function normally in life and in the kinds of jobs available. What kinds of tests show that they are? Whatever the tests, they are just symptoms or correlates of what we really want. Figuring out the most reliable correlates is a complex scientific task. Moreover, adopting a measure can have its own effects; for example, schools "teach to the test," which can bring about better test results without achieving the goal desired. One challenge still facing EBP is to develop guidance for policymakers on how to go about formulating their goals.

So far, work has focused on a later stage in the process: on predicting whether a policy will achieve an already operationalized goal. Much effort has gone into devising ranking schemes for study designs. The designs ranked are based on John Stuart Mill's "method of difference" for establishing causal conclusions, which in the ideal situation consider two groups identical with respect to everything that affects the outcome other than the policy under test. A difference in probability of outcome between the groups shows that the policy caused the outcome in some individuals in the study. Observational studies, where groups are matched on known causal factors, are low in the rankings. RCTs come out on top. That's because masking and randomizing are

supposed to make it likely that causes not known about are equally distributed in the two groups. There are, however, many other methods available in the social and biomedical sciences that can reliably establish causal conclusions (including causal Bayes nets, process tracing, derivation from sound theory, and restricted kinds of econometric modeling). So another task facing EBP is to learn how to evaluate when these other methods produce high-quality evidence and how to integrate evidence from them into synthetic reviews.

The highly regarded ranking scheme called GRADE defines quality of evidence in two contexts: The first is for *systematic reviews*, which are supposed to synthesize the best available evidence. There, quality reflects whether study designs make it likely that the study conclusions are reliable. High-quality studies provide high confidence in policy "efficacy": that the policy worked in the study population. GRADE's second context is that of *recommendation*. Policies can be accepted or rejected—and legitimately so—for many reasons, independently of how likely they are to work in the targeted setting—such as cost, the chance of beneficial or harmful side effects, political or moral acceptability, fairness, or the need to do something.

None of the study designs mentioned provides the kinds of evidence that might be relevant to these other considerations. Also, these study designs provide only very partial evidence for "effectiveness": that the policy will work here, in this new setting. Generally, policies work for a reason, and sometimes the reasons that make a policy effective in one place may not do so in another.

These reasons can usefully be classed into two categories:

1. *Support factors:* Policy programs, no matter how carefully devised, seldom contain all that is required for them to work in this or that concrete setting. Consider Street's (UK) microfinance program, founded on the example of Poland's Fundusz Mikro and other international microfinance schemes. This program did not work for the sample group in the UK, in part, it seems, because even poor people in the UK can get loans through credit cards. Inability to get credit elsewhere is a necessary support factor that was not given sufficient attention before the UK program was set up.
2. *Causal roles:* Independently of the presence or absence of support factors, a cause that can produce an effect in one place may not play the same role in another. Educating mothers about nutrition improved infant health in Tamil Nadu but not in Bangladesh, because in Bangladesh far more often the mother-in-law, not the mother, decides who eats what, and men do the shopping.

So policies need the right support factors to be in place in the setting where they will be implemented. The beauty of RCTs is that they are supposed to distribute these factors evenly between two study groups even if they are unknown. But knowing what they are and ensuring that they are in place in the target setting can make all the difference to success there. So, too, can understanding what it is about institutional structures, habits, norms, and behaviors that allows factors to play the causal roles they do. EBP is thus increasingly focused on categorizing this kind of additional information, investigating how it can be unearthed, and evaluating what kinds of evidence are relevant to assessing it.

Finally, the major challenge facing EBP is how to switch from the usual question "How can we get policymakers to use rigorous research more?" to "What decisions do various policymakers make, how can we help them make better decisions, and how can research evidence help?"

Nancy Cartwright

See also Bayesianism, Recent Uses of; Causation in the Social Sciences; Econometrics: Methodological Issues; Experiment, Philosophy of; Experimenting Society, The; Philosophy of Expertise; Policy Application of Social Sciences

Further Readings

Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: Doing it better.* New York: NY: Oxford University Press.

Coalition for Evidence-Based Policy. (2012). Retrieved from http://coalition4evidence.org

Duflo, E. (2010, May). *Social experiments to fight poverty.* Retrieved from http://blog.ted.com/2010/05/03/social_experime

Howick, J. (2011). *The philosophy of evidence-based medicine.* Oxford, England: Wiley-Blackwell.

The International Initiative for Impact Evaluation. (2012). Retrieved from http://www.3ieimpact.org
National Research Council (in press). *Evidence for use.* Washington, DC: Author.
Pawson, R. (2006). *Evidence-based policy: A realist perspective.* London, England: Sage.
Pawson, R., & Tilley, N. (1997). *Realistic evaluation.* London, England: Sage.
World Bank. (1995). *Tamil Nadu and child nutrition: A new assessment.* Washington, DC: World Bank.
Worrall, J. (2002). What evidence in evidence-based medicine? *Philosophy of Science, 69,* S316–S330.
Worrall, J. (2007). Why there's no cause to randomize. *British Journal for the Philosophy of Science, 58,* 451–488.

Evolutionary Ethics

In the most general terms, evolutionary ethics is the study of the relationship between evolution by natural selection and moral discourse and practice. That relationship is multifaceted. Contemporary evolutionary ethics covers a range of different research programs, from traditional moral philosophy to behavioral economics to neuroscience. In its earliest stages, not long after the theory of evolution by natural selection took hold, evolutionary ethics was principally a normative investigation—that is, an investigation into what the story of human evolution reveals about how we ought to live our lives. Current research is largely non-normative. On the one hand, cognitive scientists (from fields such as psychology, sociology, anthropology, and behavioral economics) are exploring the influences of natural selection on our (so-called) moral minds. Did we, they ask, evolve to think and behave morally? On the other hand, moral philosophers are exploring the relationship between natural selection and the nature of moral properties themselves. According to one increasingly popular metaethical view, natural selection favored creatures that *believed* that some acts are, for example, objectively prohibited, but apart from that influence, there are no independent grounds for believing that any acts *really are* objectively prohibited.

This entry explores both the empirical and the philosophical approaches to evolutionary ethics, beginning with the former.

Empirical Approaches

Contemporary empirical research into the evolutionary roots of our moral minds was energized by an approach to the mind inaugurated in earnest in the 1970s: evolutionary psychology (EP). The underlying premise of EP is that, like other observable phenotypic traits (e.g., the structure of the heart or skin pigmentation), some/many/all of our current mental traits were biological solutions to adaptive problems routinely confronted by our hominid ancestors. The mental traits of most interest to EP, of course, are those that issue in behavior. So, to take some of the more notorious traits, it has been maintained that the reason why males typically prefer nubile females while females prefer males of means and status is that these preferences tended, over time, to increase individual reproductive fitness.

From this vantage, it was reasonable to speculate that there must be evolutionary reasons why humans tend to deploy moral concepts. One reason that has figured in several contemporary accounts is that thinking morally was an effective means of ensuring social cohesion. Anthropologists maintain that individual survival depended critically on social cohesion. Beginning with one's immediate family, the spheres of interdependence spread outward to relatives, neighbors, and other members of the group. Cooperating with these individuals was essential for things such as child rearing, protection, food gathering, hunting, and so on. But while cooperation is a more prudential strategy than "going it alone," unswerving cooperation has its limitations. Occasions will inevitably arise in which one can gain more by exploiting the cooperation of others than by doing one's fair share (as powerfully demonstrated by Prisoner's Dilemma—style games). This strategy of occasional exploitation, however, is unstable under the assumption that most individuals interact regularly over time. For it is thought that regular interactions would inevitably expose cheaters. And over the long run, cheaters who were exposed suffered more than cooperators, who always refrained from exploiting others. It is here that some believe moral thinking earned its spot. An individual who believed that cheating others was prohibited—was *immoral*—and who felt the corresponding motivation to abide by that belief enjoyed the long-term benefits of appearing trustworthy. In rough outline, this is the evolutionary account.

Support for the evolutionary account is plainly intuitive. But some maintain that there is a growing body of empirical support as well. If our moral minds did have an evolutionary basis, then it should be part of our genetic inheritance. And if it is part of our genetic inheritance, then we should see it develop (ontogenetically) relatively early on and across many different environments. Some developmental psychologists cite evidence that children as young as 4 and 5 years of age demonstrate a moral competence that could not have been mastered simply by observing one's environment. For example, young children distinguish between moral rules (e.g., "Don't hit") and conventional rules (e.g., "Don't throw food") despite any apparent training. Around the age of 5, children suddenly begin to measure moral worth in terms of intentions as opposed to merely consequences, a reversal that does not appear grounded in environmental cues. Such poverty-of-stimulus arguments drive some to insist that such competencies must be innate. Additionally, some anthropologists insist that all extant cultures exhibit adherence to some moral code or other, and such cultural convergence is indirect evidence of a shared genetic program.

Skeptics, on the other hand, argue that the evolutionary account is not supported by conclusive evidence. For example, some insist that successfully accounting for children's moral competence does not require positing an innate moral faculty. It may be that children merely possess a disposition to acquire social norms from their environment, perhaps ones that are backed by strong emotions. And the evidence from cultural anthropology, according to some, suggests substantial cultural *disparity* in moral outlook —not what the evolutionary account predicts. Finally, any hypothesis about innate mental faculties will eventually need to link genes, brain, and behavior, for the evolutionary account ultimately comes down to the transmission of genes whose operation reliably produced adaptive behavior. But establishing those links has been difficult in the extreme, largely because genes appear to stand in a many-to-many relationship with phenotypic traits. Moreover, as all sides agree, there is no localizable brain system devoted to morality, so the task requires identifying a set of genes responsible for a set of brain systems, none of which exclusively underwrite morality.

We are, at any rate, in the early stages of this research. No one accepts the idea that moral development in individual humans is as biologically routine as, say, the development of sex organs. By the same token, no one accepts the idea that moral development does not require the deployment of multiple innate brain systems (no one, after all, learns to grow an amygdala).

Philosophical Approaches

The second broad area of inquiry concerns more traditional moral philosophical questions. In general, the aim is to understand the relationship, if any, between the principles constitutive of natural selection and the way we *ought* to live our lives. Perhaps the most tempting suggestion—and one that persists in some quarters—is that "survival of the fittest" is a moral imperative: Each individual should be responsible for each individual. We might be compelled to preserve cooperative arrangements, but only because these are good for each. We have no moral responsibility to sacrifice our own well-being for the sake of others, for intervening on behalf of the needy is interfering with what Herbert Spencer called a "large, far-seeing benevolence." According to Spencer, Darwinian evolution—if allowed to proceed unchecked —invariably leads to social harmony. Nature, it was thought, is self-correcting. (To the extent that advocates of laissez-faire capitalism sought moral justification for their view, it is hard to imagine a better fit.) This view came to be known as Social Darwinism.

Despite its early promise, however, Social Darwinism confronts sizable obstacles. The first is biological. Evolution by natural selection does not favor—in the short or long run—social harmony. Allowing evolution to "proceed unchecked" no more guarantees a species of socially harmonious creatures than it guarantees a species of disharmonious leaf eaters. The products of evolution are deeply contingent.

The second obstacle faced by Social Darwinism is conceptual and has, historically, assumed two forms. There is, first, what is sometimes called Hume's Law of "fact/value distinction," after the Scottish philosopher David Hume. Hume's Law prohibits deducing normative conclusions from exclusively non-normative premises or deducing an "ought" from an "is." In other words, claims about how things *ought* to be are never logically entailed by claims (all of which are) about how things *are*. Therefore, a complete specification of our species' evolutionary history (a fact) does not logically necessitate that we ought to do or to value one thing rather than another (an

ought). Second, the early-20th-century British philosopher G. E. Moore offered a distinct but related criticism in line with the fact/value distinction. According to Moore, any attempt to identify "the Good" (or "the Moral") with any other property, biological or otherwise, is bound to fail. It commits what Moore called the *naturalistic fallacy.* For example, suppose it is proposed that all, and only, acts that promote social harmony are good. Then, on this view, the Good *just is* whatever promotes social harmony. But this can't be correct, argued Moore, for if the Good is nothing other than whatever promotes social harmony, then for any act that promotes social harmony, we cannot intelligibly ask, "Is it good?" But we can. Such a question is indeed open. This, however, indicates that there is no necessary relation between whatever promotes social harmony and the Good. For "Good" and "socially harmonious" are not synonymous and hence do not form an analytic sentence or a logical tautology. Moore was thus led to the view that the Good is an unanalyzable, nonnatural property.

While any complete survey of Hume's and Moore's challenges to Social Darwinism will issue important qualifications (e.g., Moore failed to recognize that some identities [e.g., water and H_2O] may not be synonymous), the prevailing obstacles remain. Social Darwinism, at least for the time being, has been abandoned. But this does not mean that the philosophical interest in the relationship between evolution and ethics has in turn dried up.

Beginning in the 1970s, authors began to turn the relationship between evolution and ethics on its head. Instead of maintaining that evolution *justifies* a given moral system (as Social Darwinists maintained), contemporary authors argue that evolution *explains* our tendency to believe in moral systems generally. This latter hypothesis is put forward as a debunking thesis: We believe in right and wrong because it served a biological purpose during the long period of hominid development—not because our moral beliefs "track" independently existing moral properties. In other words, moral judgments are not the result of perceptual activity; we don't perceive "not-to-be doneness" out there in the world. Rather, they are the result of internal regulation. Our minds were designed to regulate and preserve (inter alia) social bonds. And one effective means of doing so involves the postulation of moral obligations and prohibitions (recall our earlier discussion of how our moral minds might have evolved). But this empirical hypothesis allegedly yields an important epistemological implication: Belief in the objective status of moral properties is unjustified.

The argument is sometimes expressed this way: If we have evidence that a set of beliefs has been produced by means *unrelated* to the facts that would make those beliefs true, then we ought not to assent to those beliefs. According to one proposal (see above), our moral beliefs (or, more specifically, our disposition to adopt moral beliefs) are the result of ecological pressures on early hominids, not rightness or wrongness itself. Hence, we should be agnostic about the objective status of morality.

One way of responding to this argument is by questioning the antecedent. It might be argued, for example, that while some moral beliefs are the result of ecological pressures, not all of them are. Alternatively, one might insist that morality itself consists in the very drive to social cohesion to which early hominids would have been sensitive; thus, it *was* a response to rightness and wrongness that shaped our moral minds. Another way of objecting to the argument might consist in questioning the conditional: Many of our beliefs were produced by means unrelated to the facts that would make them true (e.g., my belief that 2 + 2 = 4 was produced by the utterances of one Miss Peacock in the second grade), yet we are nonetheless justified in continuing to hold them.

Scott M. James

See also Biology and the Social Sciences; Cooperation, Cultural Evolution of; Cultural Evolution; Emotions in Economic Behavior; Evolutionary Game Theory and Sociality; Evolutionary Political Science; Evolutionary Psychology; Moral Cognitivism; Neuroethics; Sociobiology

Further Readings

Boniolo, G., & De Anna, G. (Eds.). (2009). *Evolutionary ethics and contemporary biology.* Cambridge, England: Cambridge University Press.

Carruthers, P., Laurence, S., & Stich, S. (Eds.). (2005–2008). *The innate mind* (3 vols.). Oxford, England: Oxford University Press.

James, S. M. (2011). *Introduction to evolutionary ethics.* Malden, MA: Wiley-Blackwell.

Sinnott-Armstrong, W. (Ed.). (2008). *Moral psychology: Vol. 1. The evolution of morality.* Cambridge: MIT Press.
Thomson, P. (Ed.). (1995). *Issues in evolutionary ethics*. Albany: SUNY Press.

Evolutionary Game Theory and Sociality

This entry introduces the novel notion of evolutionary game theory and its recent uses in explaining, naturalistically, the reality of social conventions as well as sociality in general. It explains what game theory aims at doing and how its borrowing conceptual tools from evolutionary theory in biology can help explain human social conventions and coordinated action, including social phenomena such as mutual trust. It focuses on explaining the main elements of such a naturalistic account, namely, the notions of conventions, equilibrium, reciprocity, and emergent phenomena, and ends by giving an example of how to naturalize along those lines a political theory of justice and egalitarianism.

Naturalistic Explanation

Evolutionary game theory has nothing to offer those who follow Immanuel Kant in seeking metaphysical foundations for human morality. Its role lies in providing a tool for those who approach moral issues from the naturalistic perspective of David Hume. John Mackie's *Inventing Right and Wrong* is perhaps the most vigorous modern exposition of this position. After a fierce critique of standard metaphysical arguments, he argues that alternatives need to be based on anthropological data analyzed using game theory as a conceptual framework.

Conventions

The idea that a social norm should be identified with a Humean convention was pursued in modern times by Thomas Schelling and then David Lewis, before being taken up by game theorists. Schelling speaks of *focal points* rather than *conventions*. Robert Sugden usually says *salient* rather than *focal*. Cristina Bicchieri and Herbert Gintis offer differing psychological definitions of the term *social norm* that would exclude many social conventions. Ken Binmore and Brian Skyrms hijack the traditional term *social contract* to refer to the collection of all social conventions or norms that regulate a particular society.

However, in spite of their differences, naturalists all agree that social conventions govern a wide spectrum of social behavior, from the trivial to the profound. Their range extends from the arcane table manners we employ at formal dinner parties to the significance we attach to the green pieces of paper we carry around in our wallets bearing pictures of past presidents (in the United States), from the side of the road on which we drive to the meaning of the words in the language we speak, from the amount people tip in restaurants to the circumstances under which we are ready to submit ourselves to the authority of others, from the order in which we arrange words in dictionaries to the criteria we apply when making judgments about fairness.

Coordination Problems

From the perspective of game theory, human social life consists largely of the play of a succession of coordination games.

Who goes through that door first? Who gives way to whom when cars are maneuvering in heavy traffic? Whose turn is it to wash the dishes tonight?

When interacting with people from our own culture, we usually solve such coordination problems so effortlessly that we do not even think of them as problems. It is only after a coordination problem has been modeled as a game that it becomes apparent that social criteria are needed to determine which of the various different ways of solving the game is to be used.

For example, most countries solve the Driving Game by operating the convention of driving on the right, but in other countries the convention is to drive on the left. Hume saw that the same considerations that apply to such humdrum coordination problems might also apply to the grand problems of running a whole society. If he is right, then our social contracts reduce to bunches of conventions that together isolate one of the many ways of solving our society's game of life.

Why Do Social Contracts Work at All?

The players of a society's game of life are its citizens—including kings, the police, and all other authority figures. Its rules are all relevant constraints that we are physically unable to break.

When we seek to paint on such a broad canvas, an immediate problem arises. What is the "cement" that holds societies together? It cannot be the law or the constitution. They are just words on pieces of paper. Nor is it the officers of the state. They are just people like you or me. Is it our sense of moral obligation? If so, what is its source? Why does it command authority?

Hume's answer is that stable social contracts need no cement. They hold together like a masonry arch. Each stone supports and is supported by its neighbors, without any need for cement. Game theorists express this Humean idea by saying that the mutual understandings built into a stable social contract succeed in coordinating our behavior on an *equilibrium* in our game of life.

Equilibrium

A *game* is any situation in which people or animals interact. The plans of action of the players are called *strategies*. A *Nash equilibrium* is any profile of strategies—one for each player—in which each player's strategy is a best reply to the strategies of the others. Nash equilibria are of interest for two reasons:

1. If it is possible to single out the rational solution of a game, it must be a Nash equilibrium—if Adam knows that Eve is rational, he would be stupid not to make the best reply to what he knows is her rational choice. *This is the basic principle of rational game theory.*
2. An evolutionary process that adjusts the players' strategy choices in the direction of increasing fitness can only stop when it reaches a Nash equilibrium. *This is the basic principle of evolutionary game theory.*

Because simple *evolutionary* processes stop working at a Nash equilibrium, biologists say that Nash equilibria are "evolutionarily stable" (usually ignoring the fact that their formal definition includes further small print). At an evolutionarily stable outcome, each relevant locus on a chromosome is occupied by the gene with maximal fitness. Since a gene is just a molecule, it cannot *choose* to maximize its fitness, but evolution makes it seem as though it had. This is a valuable insight, because it allows biologists to use the rational interpretation of an equilibrium to predict the outcome of an evolutionary process, without following each complicated twist and turn that the process might take or attributing cognitive abilities to animals (or humans) that are beyond their capacity.

Note that evolutionary game theory says nothing whatever about the psychological mechanisms involved in operating an equilibrium. If someone treats us unfairly, we may get angry and respond negatively. Game theory can explain that fairness norms would not survive if deviations were not punished, but it offers no insight into how it feels to be angry or what stories we tell ourselves to justify our not turning the other cheek. That is to say, game theory can only offer ultimate explanations of our behavior. However, the fact that proximate explanations are also available (and are much more interesting in many situations) does not invalidate the explanations offered at the ultimate level.

Cultural Evolution

For many applications, it is necessary to appeal to cultural or social evolution rather than biological evolution. Imitation, education, and individual trial-and-error learning then substitute for genetic replication. J. McKenzie Alexander and Skyrms have written books for philosophers that explain how such cultural processes can coordinate human behavior on the kind of focal points considered by Schelling.

Reciprocity

Confucius was once asked to summarize the secret of human sociality in a single word. He supposedly replied *reciprocity.*

Game theorists rediscovered his secret when characterizing the outcomes that can be supported as Nash equilibria in repeated games. The result of this work is known as the *folk theorem,* since it was formulated independently by several game theorists in the 1950s. It says that no social cement is necessary to persuade a collection of Mr. Hydes to cooperate like Dr. Jekylls. It is only necessary that they be patient and know that they are to interact together for an indefinite period in the future. The rest can then be left to their enlightened self-interest, provided that they can all monitor each other's past behavior without too much effort—as, for example, when humans were all members of small hunter-gatherer communities.

Repeated games satisfying the conditions of the folk theorem normally have an infinity of Nash equilibria, both efficient and inefficient. They therefore pose the equilibrium selection problem in an acute form. Our current convention might perhaps require us to operate in an inefficient equilibrium, but the folk theorem tells us that *any* efficient outcome that we all prefer to the status quo can also be sustained as a Nash equilibrium—and hence we could all be made better off by changing our convention. Robert Axelrod's computer simulations go on to suggest that we should expect evolution to organize such a reform spontaneously on our behalf if given long enough to operate in a stable environment.

Space limitations do not allow a discussion here of the difficulties that arise when seeking to see how far the folk theorem extends to a modern society in which each player's past history of play is often anything but an open book. On the other hand, the theorem does provide the beginnings of an explanation of how cooperation is possible in societies without any need to invent some kind of moral cement.

Emergent Phenomena

Game-theoretic models of sociality have no primitives for notions such as authority, blame, courtesy, duty, envy, friendship, guilt, honor, integrity, justice, loyalty, modesty, ownership, pride, reputation, status, trust, virtue, and the like. They are seen as *emergent* phenomena that describe aspects of whatever equilibrium may have evolved in a society's repeated game of life. Binmore's efforts in this area have been directed at explaining how the egalitarian intuitions of John Rawls's political philosophy developed in his classic *Theory of Justice* can be *naturalized* in this way. How is this done? Space allows only the briefest discussion of the much simpler notion of *trust*.

Alice delivers a service to Bob, trusting him to reciprocate by making a payment in return. But why should he pay up if nothing will happen to him if he doesn't? If Alice predicts this response, why should she deliver the service? However, people mostly do pay their bills. When asked why, they usually say that they have a duty to pay and that they value their reputation for honesty. Game theorists agree that this is a good proximate description of how our social contract works, but it needs to be supplemented with the *ultimate* explanation that the indefinitely repeated version of the game played by Alice and Bob has an equilibrium in which Alice always delivers and Bob always pays. We explain this *equilibrium* in real life by saying that Bob cannot afford to lose his reputation for honesty by cheating on Alice, because she will then reciprocate by refusing to provide any service to him in the future. In practice, Alice will usually be someone new, but the same equilibrium works just as well, because nobody will be any more ready than Alice to trade with someone with a reputation for not paying.

Conclusion

Criticism, of various sorts, has been directed at the naturalistic approach outlined here. Apart from the traditional philosophical objections, there are new complaints from behavioral economists, who see no need for a folk theorem at all.

Ken Binmore

See also Conventions, Logic of; Cooperation, Cultural Evolution of; Cooperation/Coordination; Cultural Evolution; Emergence; Game-Theoretic Modeling; Naturalism in Social Science; Social Contract Theories; Social Conventions; Social Norms; Social Rules; Trust, Social

Further Readings

Alexander, J. M. (2008). *The structural evolution of morality.* Cambridge, England: Cambridge University Press.

Axelrod, R. (1984). *The evolution of cooperation.* New York, NY: Basic Books.

Bicchieri, C. (1990). Norms of cooperation. *Ethics, 100,* 838–861.

Binmore, K. (2005). *Natural justice.* New York, NY: Oxford University Press.

Gintis, H. (2009). *Bounds of reason: Game theory and the unification of the behavioral sciences.* Princeton, NJ: Princeton University Press.

Lewis, D. (1969). *Conventions: A philosophical study.* Cambridge, MA: Harvard University Press.

Mackie, J. (1977). *Ethics, inventing right and wrong.* London, England: Penguin Books.

Mailath, G., & Samuelson, L. (2006). *Repeated games and reputations: Long-run relationships.* New York, NY: Oxford University Press.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives, 14,* 137–158.

Schelling, T. (1960). *The strategy of conflict.* Cambridge, MA: Harvard University Press.

Skyrms, B. (1996). *Evolution of the social contract*. Cambridge, England: Cambridge University Press.
Skyrms, B. (2003). *The stag hunt and the evolution of the social structure*. Cambridge, England: Cambridge University Press.
Sugden, R. (1986). *The economics of rights, cooperation and welfare*. Oxford, England: Blackwell.

Evolutionary Political Science

Evolutionary political science involves the application of principles from evolutionary biology and evolutionary psychology to concepts and problems drawn from political science. Evolutionary political science seeks to understand and interpret complex social and political behavior in light of the reproductive advantage offered by particular strategies. These theories explain aspects of behavior that involve universal characteristics, or the ways in which humans are similar. Thus, these theories remain distinct from models that focus on exploring the nature of individual difference. Most examinations of unique facets of human political behavior employ methods from behavior genetics. Evolutionary and genetic perspectives, while completely complementary, are not identical.

This entry explicates the basis of these evolutionary models and then discusses some areas of application to political science.

Evolutionary Models

Evolutionary models examine the ways in which humans and other animals adapted to repeated challenges over their ancestral past. Only those problems that occurred repeatedly and exerted an influence on reproductive success are subject to evolutionary pressure. From an evolutionary perspective, natural selection operates over time and across millions of people in such a way that the marginally more successful strategies proliferate in a population, while those that are less successful diminish in prevalence over time. As a result, evolutionary theory suggests that individuals are born into the world with certain inherent biological and psychological mechanisms. These mechanisms have a general content but often remain specific to a particular domain. This means that individuals will not respond in the same way across situations but rather will know to avoid predators but approach family members and other allies. As a result, the physical and psychological systems that often utilize the same responses, such as anger, must remain flexible enough to operate differently across various circumstances. Moreover, repeated problems have unique environmental cues or triggers to signal their appearance. For example, war represented a repeated challenge that required the evolution of mechanisms to recognize, develop, and maintain coalitions so that individuals could both work together to fight enemies as well as distinguish friends from foes. In this way, evolutionary models remain fundamentally contingent on environmental inputs.

Evolutionary models cannot explain everything. They would not expect psychological mechanisms to exist for problems that did not appear repeatedly in the ancestral past or whose successful resolution did not have an impact on reproductive success. Furthermore, evolutionary systems are backward looking: Strategies that proved successful in the past may not remain optimally adaptive in the current environment, but this does not invalidate the model itself. For example, efficient fat storage in the past would have protected individuals from the ravages of famine but could lead to obesity and heart disease in the modern world of sedentary lifestyles and fast food. However, environmental cues that triggered particular responses in the past, for example, territorial invasion causing desire for vengeance, may automatically trigger similar responses in the current environment.

Evolutionary models can offer predictive and testable hypotheses, but the nature of that prediction can differ from that offered by more established models in political science. Many political models strive for prediction that focuses on timing, attempting to specify when a particular event will happen. Others concentrate more on whether a certain event or action will occur. By contrast, evolutionary models focus on specifying and predicting the patterns of behaviors and events that might take place under particular environmental circumstances.

Applications to Political Behavior

There are many areas where evolutionary models can provide novel insight into enduring problems in political behavior. Evolutionary models offer a unique approach to the study of coalitional behavior, including problems related to labor recruitment, resource sharing, and reputation and status.

Evolutionary models can also inform understanding of the nature of leadership and followership. In addition, an evolutionary perspective supplies an alternative explanation for the nature of threats, conflict, overconfidence, and war.

Additional enduring problems in political science can benefit from applying an evolutionary lens. For example, evolution can help observers gain better traction on how political groups such as states, organizations, and institutions are perceived and represented by both individuals and other states. In particular, humans evolved in small hunter-gatherer groups and would not be expected to formulate automatic representations of large groups such as the modern nation-state in that context. As a result, humans likely represent groups as a kind of special category of individuals, infusing attributes of both unitary actor and intentionality to structures that in reality possess neither. An extension on this understanding could explain how coalitional action, such as alliance formation, maintenance, and dissolution, or institutional reform, might be facilitated by state actors or other organizations. Leaders and other coalitional entrepreneurs, for example, likely seek opportunities to strategically manipulate the coalitionary environment by using devices such as emotional outrage to prime and lubricate collective action, such as war. Such action may elicit instinctual responses from followers, although it may be in the service of individual leadership goals in a modern environment.

Finally, evolutionary theory suggests a new foundation for feminist theory in political science as well, since many of the behaviors explained by such models, for example, aggression, display demonstrable sexual dimorphism. For example, some interesting work has shown a link between anger, physical formidability, and aggression. Specifically, male, but not female, coalitional size and physical strength predicts support for aggressive foreign policy intervention.

In short, evolutionary models provide powerful and parsimonious models for explanation and prediction in political science.

Rose McDermott

See also Biology and the Social Sciences; Evolutionary Ethics; Evolutionary Game Theory and Sociality; Evolutionary Psychology; Feminism: Schools of Thought; Genetic Indeterminism of Social Action; Neuroscience and Politics

Further Readings

Hatemi, P. K., & McDermott, R. (2011). *Man is by nature a political animal: Evolution, biology and politics*. Chicago, IL: University of Chicago Press.

Lopez, A. C., McDermott, R., & Petersen, M. B. (2011). States in mind: Evolution, coalitional psychology, and international politics. *International Security, 36*(2), 48–83.

Sell, A., Cosmides, L., & Tooby, J. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences, 106*(35), 15073–15078.

Thayer, B. (2004). *Darwin and international relations: On the evolutionary origins of war and ethnic conflict*. Lexington: University of Kentucky Press.

Evolutionary Psychology

While it has long been recognized that humans are the product of evolution, progress in applying evolutionary principles to the study of the mind has been slow. Much of this delay has come from disagreements about ontology: What are the elements of which the mind is composed, and how are these shaped by evolutionary processes? This is the central question that evolutionary psychology seeks to ask. By decomposing the mind into its underlying functional processes, evolutionary psychology stands to shed light on many philosophical questions about the nature of the mind.

The research strategy of evolutionary psychology is to use empirical methods, guided by evolutionary considerations, to attempt to decompose the mind into its underlying information-processing components or adaptations. The mind is a complex organ that is likely to be composed of multiple, distinct adaptations rather than a single one, and the nature and arrangement of these adaptations must be discovered empirically. Since the field's origins in the late 1980s and early 1990s, evolutionary psychologists have developed a body of theories and methods that have been applied to diverse aspects of the study of the mind, from neuroscience and development to culture and social behavior.

Evolutionary Psychological Theory

Adaptations

Evolutionary psychology begins by treating psychology and the other social sciences as branches

of biology. Because brains and minds (used synonymously here) are parts of an organism's phenotype, the causes of their organization must ultimately be biological, and evolutionary, in origin. The design features of brains and their underlying mechanisms—mechanisms of perception, emotion, cognition, and learning—evolve because of the effects these mechanisms have on organismic survival and reproduction, or fitness. Mechanisms that evolve in this way are known as *adaptations*. Crucially, the fitness effects that shape mechanisms are not abstract but arise because of interactions between the organism and the environment. The set of environments that has shaped the evolution of a given brain mechanism is known as its *environment of evolutionary adaptedness* (EEA).

EEAs and Domains

Because the brain of any organism is an evolutionary mosaic, composed of multiple adaptations with distinct evolutionary histories (some ancient in origin, some more recent), there is no single EEA for an entire organism. For example, human adaptations for a meat-rich diet, which are relatively recent in origin, have a briefer evolutionary history than mechanisms in the retina for detecting light, which are ancient. Moreover, adaptations differ with regard to which aspects of the environment they interact with to influence fitness. This subset of features is referred to as an adaptation's *domain,* and the domains of adaptations differ. For example, the domain of our visual system (properties of visual light and the manner in which it reflects from objects) is distinct from the domain of our auditory system (properties of sound waves and how they reflect from objects). Because most psychological adaptations within the mind are designed to process information coming from elsewhere in the brain rather than directly from the external environment, their domains include internal information. For example, search-and-recall mechanisms in human memory have evolved to interface with human memory stores, so they must contain adaptations to the properties of memory storage. Finally, in a social species such as humans, the domain of psychological adaptations may also include the social environment. For example, adaptations for cultural learning are adapted to the environment of cultural information transmitted by other humans.

Phylogeny and Descent With Modification

Adaptations do not arise de novo, with no evolutionary history. Instead, they arise through variations on previously existing structures, a process Charles Darwin referred to as "descent with modification." This has important consequences for the study of psychological adaptations. It means that every adaptation contains elements, inherited from prior structures, that are not the result of selection for the adaptation's current function. It also means that many adaptations possess hierarchically nested design features, some ancient and some recent. For example, human mating psychology may contain human-specific features nested within a more primatewide or mammalwide mating psychology. Finally, it means that adaptations are distributed phylogenetically across taxa, such that an adaptation in one organism can be evolutionarily related, or *homologous,* to a modified version of that adaptation in another organism. This means that "human nature" must be considered largely as a set of modifications of more ancient mechanisms rather than as having arisen entirely since the chimp–human common ancestor.

Mechanisms

Evolutionary psychologists typically take a functionalist approach to thinking about psychological mechanisms, in which mechanisms take inputs (i.e., information from the environment or internally), operate on those inputs, and send outputs to other brain systems. Any aspect of a mechanism's input, output, or processing procedures that has been shaped because of its impact on fitness in ancestral environments is called a *design feature*. While this approach was originally applied to perceptual mechanisms, it can be and has been expanded to cover virtually all aspects of brain functioning, from reasoning, to decision making, to emotion, to mechanisms of learning and brain development.

Modularity

The concept of an evolved psychological adaptation, central to evolutionary psychological approaches to the mind, has sometimes been taken as synonymous with the concept of a "module" as it is used in psychology and philosophy of mind. In this traditional conception, modules are held to have

several properties, including encapsulation (inability to be influenced by conscious thought), innateness (identical development across individuals, without learning), automaticity (reflex-like operation independent of other brain processes), and fast, rather than slow, operation. While this concept of modularity might usefully describe some evolved mental adaptations, it is important to realize that not all adaptations are likely to be modular in this sense. For example, conscious processes, while typically held to be nonmodular, have almost certainly been shaped by natural selection, so the concept of a psychological adaptation should not exclude interaction with consciousness by definition. Similarly, while speed is a property that might be associated with some evolved processes, such as responses to immediate danger, other mental adaptations might be expected to operate over relatively long timescales, for example, processes of bond formation between mates. Instead of mental adaptations being identifiable by a fixed checklist of modular features, such as automaticity, encapsulation, and developmental canalization, it is likely that mental adaptations vary in their properties depending on the functions they have been selected to carry out, with only some corresponding to the traditional conception of cognitive modules.

Empirical Research in Evolutionary Psychology

In principle, all functional aspects of mental processing must be the result of adaptations interacting with the world, including developmental adaptations such as mechanisms of learning and plasticity. Therefore, an evolutionary psychological approach can be applied to any area of human psychology and behavior. In practice, early evolutionary psychological research was initially focused on specific topics such as mate choice and reasoning about social contracts. However, it has since expanded to include diverse areas of psychology and the social sciences, including philosophy, economics, literature, history, and anthropology.

Perception

Perceptual mechanisms deliver information from the outside world into the brain, and natural selection shapes these mechanisms because of the effects they have on survival and reproduction. This can explain both what organisms are able to perceive and what their brains do with the perceived information. For example, bees perceive high-energy ultraviolet light because flowers use it to signal the location of nectar, and deep-sea animals are tuned to be maximally sensitive to blue wavelengths that penetrate into deep water. Often, perception is not "veridical," as when we perceive an object that reflects a very different light spectrum in sunlight and in the shade as being the "same" color. This phenomenon of "color constancy" is thought to be an adaptation for tracking objects under different lighting conditions. Even within the perceptual array, what we are tuned to notice most easily also depends on fitness considerations, for example, we more easily notice the movements of animate than inanimate objects in a scene.

Emotion

Emotions and other affective states are the result of processes that add value to objects and situations, shaping organisms' motivations with respect to those things because of their effects on fitness. For example, disgust appears to be an adaptation designed to prevent organisms from ingesting pathogenic substances and is caused by perceptual cues that an object may be a source of contagion. Fear, on the other hand, is designed to cause avoidance of injury and is triggered either by harmful agents, such as predators, or harmful objects, such as weapons or cliffs. Other emotions cause us to approach objects or situations that may have fitness benefits: Hunger increases attraction to food, and sexual desire increases attraction to mates. Even social emotions have design features that make sense in fitness terms. For example, anger, or punitive sentiment, appears designed to prevent exploitation by others in social situations.

Reasoning, Judgment, and Decision Making

While some researchers have held that only "lower-level" brain mechanisms such as perception and emotion have been shaped by natural selection, it is clear that our higher-level faculties of thought, including conscious choice itself, must also be products of evolution. Whereas approaches to cognition in philosophy of mind often take standards such as truth and rationality as the normative benchmarks against which cognition should be measured, an evolutionary psychological approach assumes that

fitness is the primary factor that shapes cognitive mechanisms. When fitness considerations and rationality diverge, apparent irrationalities may result. Researchers in judgment and decision making have documented a host of such apparent rationalities that may have evolutionary explanations. For example, the "endowment effect"—the tendency to place a higher value on a good that one owns than on an identical good that one does not own—may appear economically irrational, but it could be a fitness-promoting strategy if owners have an advantage over nonowners in contests, which appears to be the case across many animal species.

Learning and Development

Mechanisms that shape organismic development, including mechanisms of learning, must also be the products of natural selection and therefore evolve because of their effects on fitness. Behaviorist psychologists originally held that all learning could be explained by two mechanisms: operant and classical conditioning. However, evolutionary research suggests that the landscape of learning may be much more complex. For example, among primates, humans are especially good at learning from observing others, and this likely requires specialized social learning abilities, such as the ability to make inferences about others' goals. Human infants do not "blindly" imitate but imitate what they infer another is trying to do, especially when being explicitly taught. Some classes of information—especially fitness-relevant information such as danger—are more easily learned than others, a phenomenon known as *prepared learning*. And, finally, development proceeds at different rates across domains, with some skills, such as understanding the mechanical properties of objects and attending to the mental states of others, learned early and with little effort and others, such as mathematics and reading, developing only later and under special cultural circumstances. This suggests the presence of adaptations for the early development of certain fitness-relevant skills in childhood.

Culture

The study of culture is increasingly coming to be a part of evolutionary psychology, with the realization that culture cannot be simply the result of general-purpose learning mechanisms. The kind of cumulative cultural evolution seen in humans is not present in other species. Evolutionary theorists have begun to study the specialized mechanisms that evolve because of the benefits cultural transmission provides, such as biases for conforming to societal norms and imitating the prestigious. Whereas culture and biology have previously been seen as mutually exclusive explanations for aspects of human behavior, the realization that culture is a human adaptation, enabled by specialized mechanisms, brings cultural phenomena into the spotlight of evolutionary explanation.

Evolutionary Psychology and Philosophy

Philosophical questions hinging on the nature of mind may be illuminated by an evolutionary approach. For example, an evolutionary approach to phenomenology may help answer ancient philosophical questions about perception, such as why experiences feel the way they do (e.g., why sugar tastes good and sand does not). Evolutionary approaches can be applied to questions in epistemology, as in the field of "naturalized epistemology," as well as semantics and pragmatics. Most recently, work on the evolution of cooperation and sociality has begun to be applied to questions in moral philosophy, with empirical work beginning to shed light on questions such as which aspects of morality are universal, and which are variable, across societies. Finally, an evolutionary approach that takes culture as a central part of human nature may speak to issues in social constructivism, including a better understanding of social norms, how they evolve, and how they influence human thought and behavior.

H. Clark Barrett

See also Biology and the Social Sciences; Cooperation, Cultural Evolution of; Cultural Evolution; Developmental Psychology; Evolutionary Ethics; Evolutionary Political Science; Human Cultural Niche Construction and the Social Sciences; Modularity of the Mind; Primatology and Social Science Applications; Sociobiology

Further Readings

Barkow, J., Cosmides, L., & Tooby, J. (Eds.). (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York, NY: Oxford University Press.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review, 113,* 628–647.

Buss, D. M. (Ed.). (2005). *The handbook of evolutionary psychology.* New York, NY: Wiley.

Pinker, S. (1997). *How the mind works.* New York, NY: W. W. Norton.

Richerson, P., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution.* Chicago, IL: University of Chicago Press.

Wright, R. (1994). *The moral animal: Why we are the way we are—The new science of evolutionary psychology.* New York, NY: Vintage.

Existential Phenomenology and the Social Sciences

Existential phenomenology is often treated as the application of the German philosopher Edmund Husserl's general phenomenological method directly to the issue of human existence. Major figures in the phenomenological tradition after Husserl for whom concrete human existence is a focal point include Martin Heidegger, Jean-Paul Sartre, and Maurice Merleau-Ponty. For some readers, the movement from Husserl's "transcendental" concerns (i.e., seeking the absolute conditions of possibility for any "phenomena" whatsoever) to the more specific focus in these subsequent thinkers on the phenomenon of historically situated and embodied human existence marks a significant shift in the phenomenological movement. However, there are strong existential characteristics already present in Husserl's thought itself, and these are brought out well by highlighting his view of science.

Of particular note in the context of this encyclopedia is that Husserl accords a key role to the social and human sciences (*Geisteswissenschaften*) as distinctive sciences of human subjectivity. This entry uses Husserl's notion of a "crisis" in science as a way of highlighting the existential aspect of his phenomenology, outlining the particular contribution he thinks phenomenological philosophy makes to the social sciences and the subsequent influence of this view.

Husserl and the "Crisis" of the Sciences

Husserl's early analysis of calculative rationality in the 1891 work *Philosophy of Arithmetic* gives a formal sense of what can go wrong with scientific thinking. In this text, Husserl describes calculation as the human capacity to use symbolic number concepts and to engage in complex, higher-level operations without grasping how the concepts arose in the first place or possessing an "authentic" understanding of the operations within which these concepts are employed. This ability to calculate is for Husserl a significant aspect of our intellect; without this power to calculate, the human mind would be capable of little arithmetical progress. Moreover, this symbolic or "inauthentic" thought of calculation tends to be very efficient: It allows for the greatest result from the least amount of effort and hence is representative of the principle of "economy of thought" put forward by earlier thinkers such as the philosopher Richard Avenarius and the physicist-philosopher Ernst Mach, which figured significantly in conceptions of science at the end of the 19th century. However, despite playing a central role in the rapid progress that arithmetical science (and science in general) is able to make, there is a negative aspect to this power of calculation. Though economical, calculation also contains a type of "blindness" about itself. Paradoxically, it allows for a powerful "thinking" without much genuine thought. The "thoughtless thinking" of calculation, while not wholly negative for Husserl, can "get out of control" and devolve into a mere technique. What Husserl describes in one of his final works, *The Crisis of the European Sciences and Transcendental Phenomenology* (1935/1936), is precisely a full-blown critique of how natural science, human/social science, and philosophy itself have fallen into various types of "thoughtless" thinking that allow them to operate, often with great technical competence, but without genuine understanding of themselves and of the origin of their own operations.

That there is some sort of "crisis" of blindness in the natural sciences (*Naturwissenschaften*) may seem surprising. They appear to function well in their task of delivering a proper description of nature, at providing and ordering the empirical facts produced by observation. At a higher level, the natural sciences account for these facts by "discovering" essential laws of causality. These causal laws are binding in all circumstances, hence the claim of the natural sciences to both necessity and universality. The natural sciences achieve this status by excluding subjective elements in their determination of nature

and its causes. How a human subject might "feel" about a particular aspect of nature, for example, is not relevant to these sciences. They seek to describe nature as it is "in itself," regardless of the subject. This objectivity determines the procedure of the natural sciences. Their method and goal of understanding the world as nature is linked to the world viewed as a particular type of object.

In what ways, then, are the natural sciences in a crisis for Husserl? Put simply, they lose sight of their own origin, their ultimate foundation in human subjectivity. Only human subjects conduct science, and to overlook that origin or to treat the subjective origin of natural science in a purely objective way is to misconceive the starting point of the scientific project. It must be stressed that Husserl has some sympathy for this weakness of the natural sciences, for to function as natural sciences they cannot focus on their origin in human subjectivity. If they were to do so, they would become human sciences! Still, Husserl does critique the concealment of that subjective origin: The culpability of the natural sciences is not in their focusing on nature but in focusing on nature in such a way that the origin of this very subjective interest in nature is itself obscured or denigrated.

It is the task of the human or social sciences (*Geisteswissenschaften*) to pay attention to human subjectivity. In approaching "*Geist*"—that is, spiritual or cultural existence—these sciences seek, on a basic level, to provide a proper description and classification of (both individual and collective) human activity. On a higher level, they also seek to explain these human acts by determining the essential regularities that characterize subjective, human activity, what Husserl calls the laws of "motivation." It is possible to call these disciplines the sciences of *subjectivity*. Of course, one of the cultural achievements of the human subjectivity is precisely "natural science"—and so a clear consequence of Husserl's phenomenology is the need for a proper social-scientific understanding of the very human project of natural science itself.

Where, then, is the "crisis" of the human sciences? For Husserl, the main danger is the domination of the natural-scientific paradigm (naturalism) within social science. Due to the tremendous advancement of knowledge, capability of prediction, and technological manipulation, the natural sciences come to be seen as the model by which all true science is to be judged. That is, only objective sciences of "fact" are viewed as true sciences, and to be a science of facts means to treat the human world in exactly the same manner as the natural scientist does the natural world. On the basis of this domination of the natural-scientific paradigm, a dissolution of the human sciences takes place. There is a "colonization" of the human sciences by the natural sciences. The human subject is said to be treated scientifically only when done so according to the paradigm of the natural sciences; hence, human existence must be treated in an objective manner, as an object, just like the objects of nature described by the natural scientist. In this form of dissolution, the human sciences retain the status of science but do so by losing sight of their true goal and field of inquiry, namely, the subject *as* acting subject. The human sciences are, in effect, taken over by the natural sciences; they lose their autonomy, and ultimately, existence or culture is reduced to an appendage of nature.

For Husserl, the human and social sciences are more fundamental than the natural sciences. The "fundamentality" of the human and social sciences has to do with their field of investigation: In studying the working, achievements, and traditions of human cultures, the human sciences ground the natural sciences since they are ultimately sciences of that human subjectivity which is the source from which spring the objectivistic natural sciences. Due to this fundamental role, Husserl seems much harsher in his critique of the failure of the "human sciences," since these sciences seem all too willing to sacrifice "their" specific task of understanding human accomplishments to an inappropriate naturalistic or objectivistic paradigm.

Subsequent Appropriations of the Centrality of Human Subjectivity

The strong reaction to naturalism, and the focus upon human subjectivity, is thematized sharply in phenomenological thinkers after Husserl. In his *Phenomenology of Perception* (1945), Maurice Merleau-Ponty shows how no merely natural-scientific account of the human body can account for the experience of our own embodied existence. For Merleau-Ponty, "perception" as the basic structure of human subjectivity can never be reduced to objectively determined causal relations or accounted for in a fundamental way through natural science.

A proper phenomenology of embodied perception, however, would help ground the physicalist accounts of natural science, and it is not surprising that Merleau-Ponty's thought is the springboard for much of contemporary work on the relationship between phenomenology and cognitive science.

Martin Heidegger shares as well many of the main contours of Husserl's turn to human subjectivity. His description of the concrete life ("being-in the-world") of the human subject (*Dasein*) in his major work *Being and Time,* published in 1927, is framed within a critique of sciences (whether natural or social) that accept without question the "being" of their object of inquiry. Moreover, he shows how even previous philosophy when it did take human subjectivity as a theme did so with unwarranted assumptions. However, the exact manner of bringing this "subjectivity" into view also shows divergences with Husserl's project, as the issue of "history" makes clear. For Husserl, there is a danger in mere "subjectivism" as a reaction to naturalism. Just as the natural scientist has the tendency to look upon everything as nature, the social scientist can go too far in reaction to this naturalism and end up seeing everything as mere historical construction.

"Historicism" is for Husserl one of the most manifest signs of an inadequately understood science of human subjectivity. By an overemphasis on the historical aspect of human existence, Husserl felt that the entire idea of science as a name for absolute, timeless values was undermined. So if the danger of the naturalistic attitude is a restriction of science (as knowledge of what is true) to sciences of facts, the danger of the "historicist" position is that science itself seems to become impossible. This was most evident for Husserl in relation to philosophy itself. The historicist would regard philosophy as being the expression of the "worldview" of a particular historical period. There have indeed been many philosophies, and there has been conflict between these philosophies. For the historicist, this conflict is rooted mainly in the pretentious claim by these philosophies to a universal validity. Since the historicist sees philosophy as merely the expression of a particular worldview, there can be no claim to absolute universality.

Such a position represents for Husserl the complete abandonment of the idea of philosophy as a rigorous science that is methodically directed toward the establishment of universal, eternal truth. To view everything as historical is just as mistaken as viewing everything as nature; historicism leads to relativism and the complete undermining of the idea of science. Historicism has human subjectivity in view, but in a manner that results ultimately in the inability to say anything true (in an absolute sense) about that subjectivity. The mistake of the historicist might be said to be that subjectivity is construed in a false "subjectivistic" manner, that is, as merely particular, or relative. For Husserl, this conceals the crucial feature of subjectivity, namely, its transcendental aspect. So while the error of the naturalistic attitude is an objectivism that obscures subjectivity, the error of the attitude of the historicist is that of an improper subjectivism that obscures the true nature of subjectivity.

Husserl certainly viewed Heidegger's description of human existence as falling into such historicism. While Heidegger's clear intent was to yield the transcendental structure of human subjectivity, his very account reveals its irreducible historical character. The debate between Husserl and Heidegger regarding a proper account of human existence and subjectivity foreshadows much that follows in terms of the impact of phenomenology on social science. While existential phenomenology provides a solid framework for a critique of solely quantitatively oriented social science with its naturalistic assumptions and naive reflection on "culture" (*Geist*) as merely what Husserl calls "second nature," the positive task of working out the proper relation of the human, social sciences both to the natural sciences and to their own foundation in (transcendental) phenomenology has been a matter of ongoing debate and continues to play a significant role in contemporary phenomenology of social science.

Philip Buckley

See also Being-in-the-World; Embodied Cognition; Explanation Versus Understanding; Hermeneutics, Phenomenology, and Meaning; Historicism; *Naturwissenschaften* Versus *Geisteswissenschaften;* Naturalism in Social Science; Phenomenological Schools of Psychology

Further Readings

Buckley, R. P. (1996). Rationality and responsibility in Husserl's and Heidegger's view of technology. *Philosophy of Technology: Proceedings of the American Catholic Philosophical Society, 70,* 121–134.

Carr, D. (1974). *Phenomenology and the problem of history.* Evanston, IL: Northwestern University Press.
Embree, L. (Ed.). (1996). *Encyclopedia of phenomenology.* Dordrecht, Netherlands: Kluwer.
Heidegger, M. (1977). *The question concerning technology and other essays* (W. Lovitt, Trans.). New York, NY: Harper & Row.
Husserl, E. (1965). Philosophy as a rigorous science. In Q. Lauer (Ed. & Trans.), *Phenomenology and the crisis of philosophy* (pp. 71–147). New York, NY: Harper-Torchbooks.
Husserl, E. (1989). *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy: Second Book. Studies in the phenomenology of constitution* (R. Rojcewicz & A. Schuwer, Trans.). Dordrecht, Netherlands: Kluwer.
Kockelmans, J., & Kisiel, T. (Eds.). (1970). *Phenomenology and the natural sciences.* Evanston, IL: Northwestern University Press.
Spiegelberg, H. (1982). *The phenomenological movement.* The Hague, Netherlands: Martinus Nijhoff.

Existential Psychology

Existential psychology is the branch of psychology that deals with the human being's relationship to the most essential life dilemmas, the so-called big questions of life.

Contents of Existential Psychology

Many psychologists, psychiatrists, and other therapists enter their chosen field because they are attracted by the pulsating and varied nature of human life. They are fascinated by the many unique ways in which human beings can unfold their lives. They love to relate to others and to help them unravel from their misery and redirect their lives in a more constructive direction.

These professionals need a body of psychological knowledge and understanding that respects their interest in specific human lives without reducing these lives to abstract categories, cause-and-effect relationships, and statistical averages. At our universities, this commitment to human uniqueness is often met by intellectual systems of categorization, diagnosis, and explanatory concepts. What is lacking here—from the perspective of existential psychology—is phenomenology and a commitment to the actual lives of human beings.

Amedeo Giorgi proposes the term *life world* as the crux of a psychology that respects the lived life of real persons. The relationships between living persons and the worlds in which they live should be the central focus. Other concepts that aim to capture what life is about are *life feeling* (the person's spontaneous sense of being alive, of being coherent, and of a right to be here), *life courage* (the person's life feeling combined with his or her determination to conquer fear and anxiety in order to carry out his or her life project), and *life energy* or vitality (the ability of the organism to survive even under difficult circumstances and to achieve longevity).

As conceived by existential psychology, the spontaneous unfolding of our human lives as depicted by these life concepts meets the basic conditions of existence, which are ontologically determined. The way each of us forms and molds our individual lives is a result of how we meet and interpret these basic existential conditions. Do we, for instance, hide ourselves and refuse to look directly at the facts of death and the other existential realities, such as, for instance, our aging, our actual life satisfaction, as well as our satisfaction with the present societal situation? Or can we look these realities squarely in the eye and learn how to relate to them constructively, thus living more openly and freely, in a more grounded and real way? Existential psychologists use the word *authenticity* to designate this genuine life, which is a possibility open to everyone.

There are different catalogs and lists surveying the basic life conditions. The most well-known is Irvin Yalom's enumeration of four basic existential conditions: (1) that we are going to die; (2) that in decisive moments, we are alone; (3) that we have the freedom to choose our life; and (4) that we struggle to create meaning in a world in which our life's meaning is not given beforehand. According to Yalom, these four basic conditions constitute a structure that is our premise, something that all human beings are born into. The four conditions set the frame and the agenda for the life of each individual. Many people would rather avoid thinking about, talking about, and relating to these basic conditions, including death, but this fact does not weaken their impact, rather the opposite.

Another theory about life's basic conditions, developed by Medard Boss with inspiration from Martin Heidegger, lists seven fundamental traits of human life: (1) human beings live in space; (2) human beings live in time; (3) human beings unfold through

their body; (4) human beings live in a shared world; (5) human beings always live in a particular mood, a certain psychological atmosphere; (6) human beings live in a historic context; and (7) human beings live with the awareness of their own death.

Other existential authors have developed similar lists and overviews of humankind's basic life conditions. As early as half a century ago, one particularly interesting exposition was formulated in the 1950s by Erich Fromm, the German American humanistic psychoanalyst and writer: All passions and strivings of humanity, he writes, are attempts to find an answer to our existence. Fromm then goes on to mention as basic points the individual's fundamental need (a) for love (relatedness), (b) for transcending oneself, (c) for developing rootedness and a feeling of being at home, (d) for finding one's identity, and (e) for finding one's orientation and meaning in life. According to Fromm, the most specific characteristics of any human being derive from the fact that our bodily functions belong to the animal kingdom, whereas our mental and social lives belong to a human, conscious world that is aware of itself.

Philosophical Roots

The philosophical roots of existential psychology consist of a variety of approaches in existential and phenomenological philosophy, most often drawn upon in an open and undogmatic fashion. Writings within existential psychology and existential therapy find their basic concepts and approaches in the phenomenology of Edmund Husserl and Maurice Merleau-Ponty and in existential philosophy of various positions, such as those of Søren Kierkegaard, Martin Heidegger, Jean-Paul Sartre, Karl Jaspers, Gabriel Marcel, Martin Buber, Paul Tillich, and others. These philosophical authors are not bound together by any agreement as to the answers to people's existential questions (e.g., they diverge on answers of yes or no to religion and on questions concerning political commitment in social and world problems). However, they are bound together by emphasizing the priority of *existential* questions as such—stressing that human *existence* is prior to human essence, assuming that the latter has any meaning at all, which is something disputable in existentialism. They are also united in their conviction that each of us, as primary an existence, in a certain sense lives in his or her own world and that each of us has a life task of coming to terms with his or her basic existential concerns.

Existential Therapy and Other Applications

The main application of existential psychology is in existential therapy and counseling. Recently, other applications have emerged in coaching, education, philosophical counseling, and management studies.

In therapy, existential views now form a part of several schools. The following main points refer to existential therapy in its pure form, as developed by Ludwig Binswanger and Medard Boss in Switzerland; Viktor Frankl in Austria; Ronald Laing, Emmy van Deurzen, and Ernesto Spinelli in England; and Irvin Yalom and Rollo May in the United States. Existential therapy is characterized by the following salient features:

1. The therapy consistently uses *phenomenological* methods in queries and conversations. Instead of cause-and-effect questions, the therapist might ask, "Try to describe to me what your life is like at the moment, as specifically and in as much detail as possible." Through such detailed descriptions, the therapist gradually unfolds the person's life experience so that it manifests itself clearly in the space established between client and therapist.
2. The therapeutic dialogue is based on a direct, personal relation between client and therapist, not a transference relation. The emphasis is on dialogue between equals, characterized by mutual respect. Collaborative examination of the client's situation is encouraged, carried by a common curiosity.
3. The therapeutic dialogue in existential therapy examines the connections between everyday experiences and the basic existential conditions. In some cases, these connections are self-evident, as in an acute experience of crisis. In other cases, the possible connections between an everyday occurrence and a basic category are explored in a more meditative and reflexive way in establishing a so-called ontic-ontological connection.
4. In existential therapy, there is little emphasis on diagnosis. Also, the therapist does not normally initiate the therapy by tracing the client's developmental history. Both diagnosis and childhood history are thought to easily block or hinder the client's insight into his or her

possibilities in life. Instead, client and therapist immediately address what life looks like right now for the client in all its aspects.

5. The aim of existential therapy is often defined as enabling clients to live as richly textured lives as possible and to unfold and realize their potential in the complex modern world. Existential therapy does not have as its primary goal to remove symptoms, even though these symptoms will most often be alleviated. The most important goal is that the person—in the words of Boss—will be increasingly able to meet the world freely and openly. Individuals should always be understood in their relatedness, their *being-in-the-world*. Boss calls the ideal state of being-in-the-world "composed, joyous serenity," a state where the individual embraces the world with clarity and openness.

Comparisons With Other Schools

Existential psychology is a branch of psychology proper, that is, a field of research with concepts and theories about the world that may be validated or refuted empirically. Existential psychology distinguishes itself from other disciplines within psychology by requiring that the main focus of psychology must be *human life* and each individual's relation to life's basic conditions and most important questions. Existential psychology differs significantly from *mainstream and cognitive psychology* in its explicit focus on essential *life* questions as the most important subject matter of psychology and in its insistence on the *phenomenological* perspective. Therefore, psychology should primarily account for life as experienced from within rather than as behavior observed from without. This difference gives existential psychologists less "safe" knowledge about people's reactions in, say, traffic situations or child development stages but more understanding of the real aspirations and real-life problems of human beings.

Compared with Freudian and Jungian psychology, there are a number of similarities concerning the psychologist's in-depth interest in the lives of individual human beings. An important difference, however, is that the psychoanalytic traditions attach considerable significance to the consequences of childhood events, whereas existential psychologists and therapists are mainly concerned with an individual's present and future state and with their continuous openness to change. Furthermore, existential psychologists and therapists do not propound interpretations "behind the backs" of their clients but advocate detailed descriptions of their life situations and life perspectives.

Existential psychology is to some degree related to phenomenological psychology, humanistic psychology, positive psychology, and some of the narrative (social constructivist) trends in psychology.

Bo Jacobsen

See also Being-in-the-World; Existential Phenomenology and the Social Sciences; Explanation Versus Understanding; Gestalt Psychology; Life-World; Phenomenological Schools of Psychology; Therapy, Psychological and Philosophical Issues

Further Readings

Boss, M. (1994). *Existential foundations of medicine and psychology.* Northvale, NJ: Jason Aronson.

Fromm, E. (2005). *On being human.* New York, NY: Continuum.

Giorgi, A. (2009). *The descriptive phenomenological method in psychology.* Pittsburgh, PA: Duquesne University Press.

Jacobsen, B. (2007). *Invitation to existential psychology: A psychology for the unique human being and its applications in therapy.* Chichester, England: Wiley.

May, R., Angel, E., & Ellenberger, H. F. (Eds.). (1958). *Existence: A new dimension in psychiatry and psychology.* New York, NY: W. W. Norton.

Yalom, I. D. (1980). *Existential psychotherapy.* New York, NY: Basic Books.

Experiment, Philosophy of

A philosophy of experiment has been slow in coming, although experimentation was discussed intermittently since the 17th century, when the concept of experiment began to be recognized as a necessary component of the scientific quest. The reason why philosophy of experiment is a late bloomer has to do with the history of philosophy of science; during the first half of the 20th century, when logical positivism dominated the scene, the emphasis was principally

on theory, its logical structure and representational features, while experiment and its physical execution were generally considered transparent and unworthy of philosophical investigation—a mere generator of data turned into evidence. So strong was the domination of the logical positivists that ideas concerning experimentation by thinkers such as John Stuart Mill (1806–1873), Claude Bernard (1813–1878), Ernst Mach (1838–1916), and Pierre Duhem (1861–1916) were not pursued. But since the 1980s, there has been a growing philosophical and historical interest in the process of generating experimental knowledge.

This entry looks at recent developments. First, it presents the structure of experiment and its two stages and four phases. It then outlines some of the corresponding philosophical issues associated with the methodology of experimentation, in which nature, artifices, methods, theories, as well as skill are intricately combined in the pursuit of securing reliable experimental evidence for scientific knowledge.

Structure: Two Stages, Four Phases

Experiment is a sequential scheme of action to generate data that can be recast as evidence with respect to a certain phenomenon or theory. The data obtained are the result of some (designed) interaction between the entity under investigation and a controlled setup, so that the properties examined in the former are correlated to the known properties of the latter. Thus, essentially, two different tasks are performed in the execution of an experiment: They may be termed *preparation* and *testing*. Experimenters prepare a system, and then they test it. A "preparation" is a specified, known procedure that is reproducible. A "test" is similar to a preparation in that it has a specified procedure, but in addition, it triggers the interaction between the entity under study and the controlled setup. Importantly, a test includes a step whereby information that was previously unknown is recorded and made accessible to an observer, that is, the experimenter. This information—the data—constitutes, after a suitable reduction, experimental knowledge—the evidence that is sought. Such recording is not trivial, not only because identical tests following identical preparations need not have identical outcomes but primarily because this information constitutes, after a suitable reduction, the new physical knowledge. Within certain material limitations and ethical constraints, experimenters are free to choose preparations and perform tests—this is their prerogative. However, they are not free to choose the outcome of a test. They are bound to accept the information acquired. Thus, in terms of human intention and action, the "preparation" is active and the "test" is passive.

This characterization of experiment highlights one of its fundamental features, namely, the method of *variation*. By varying (the test) a certain group of elements or a single one of the system under study (the preparation), other elements will vary too or perhaps remain unchanged (the binding information). The fundamental rule of variation facilitates the severing of the many antecedents and their consequents. However, the character of the experimental procedure is so rich and varied that it seems impossible to capture all its features deductively. Given this complexity, the fundamental dichotomy of the experimental procedure in terms of "preparation" and "test" appears to be useful. The "preparation" comprises theoretical and practical phases: (1) laying down the theoretical framework and (2) constructing the setup and making it work. The "test" too consists of two phases: (3) observing and taking readings and, finally, (4) processing the recorded data and interpreting them.

1. *Laying down the theoretical framework:* Any attempt to create artificially certain physical conditions with a view to studying a particular aspect of nature takes place within some theoretical framework. This is the "background theory." Experimenters rely on this theory to advance the claim that is put to the test. Furthermore, instrumental theories govern the performance of the instruments intended to be used. In particular, there is the theory that underlies the setup.
2. *Constructing the setup and making it work:* Whereas the first phase of background theory provides the theoretical framework, the second phase is concerned with the actual construction of the setup and operating the apparatus involved, that is, the hardware. This phase fulfills the theoretical requirements stipulated in the initial phase; in other words, this is the nuts-and-bolts phase.

3. *Observing and taking readings:* The testing stage begins with the recording of information as the preparation stage turns into test. The data can be recorded by a variety of means—manually, which involves direct observation, as well as by automated procedures, including facilities for storing the data obtained.
4. *Processing the recorded data and interpreting them:* Once the data have been amassed, experimenters arrive at the final phase of the scheme. To conclude the inquiry, that is, to obtain the final result, experimenters analyze the data and then recast them as evidence so that they can be put to use in either confirming or refuting a theoretical claim, or laying an altogether new foundation. This is the goal of the experiment. Two processes are therefore involved in this final phase: first, the process of reduction, of analyzing the data with a view to obtaining a coherent and consistent result, and, second, the process of interpreting the result—turning it into evidence.

Experimenters assume in the preparation stage a theory that is considered correct; it underpins the experiment. This theory, the background theory, and associated theories of the instruments and the setup itself are therefore taken for granted. They are not tested by the experiment. Then there is the process of realizing these theoretical requirements in practice. Once the theory and its physical realization have been put to work in the preparation stage, the second stage begins, namely, the test: Information is allowed to flow to the recording device. And, finally, a process of reducing the recorded data and interpreting the result takes place. This is the conclusion of the experiment—the outcome as experimental evidence. Clearly, this complex scheme, designed to generate knowledge, comprises many different categories and harbors intricate philosophical difficulties.

Philosophical Issues

Roles of Experiment

Experiments play critical roles in scientific methodology. They are designed, among other things, to test theories, to help articulate theories, to explore and call for new theories, to exhibit new phenomena, to create the conditions for effects, to offer evidence of new entities, to perform measurements, and, in general, to provide the basis for scientific knowledge.

Epistemology of Experiment

What is the ground for believing in the outcome of experiment? Experiment is a complex scheme involving a spectrum of categories, from theory and material realization through practice and skillful manipulation of instruments to stabilizing the result and interpreting it. In view of this complexity, why is it that experimental evidence is considered fundamental to scientific knowledge? What are the epistemic means for distinguishing between valid experimental results and artifacts that can lead one astray? The epistemology of experiment concerns, then, the reasons for believing in experimental results as physical knowledge. Since it seems impossible to infer deductively all experimental techniques, it remains to construct a list of strategies that are neither exclusive nor exhaustive. In the final analysis, experimenters believe that they can ground the results they obtained on physical arguments, that is, arguments that include physical assumptions. The stage of preparation may be considered a set of such presuppositions, while the test stage results in consequences "inferred" by some physical process.

Tacit Knowledge

One of the difficulties inherent in the epistemology of experiment is the fact that it relies on knowledge that cannot be expressed propositionally. This tacit knowledge may be described simply as skill. Experimenters develop skills that they acquire after spending much time with the system they experiment on. They cultivate a sensitivity for the setup, which essentially indicates when the experiment is functioning properly. This intuitive knowledge cannot be rendered rational in the strict sense of the term.

Regress

In the process of constructing the setup and running the experiment, experimenters make numerous assumptions. The most common of these is that the parts of the setup and their arrangement stand up to the required specifications. To be sure, it is possible to ascertain some of these assumptions by putting them to the test. However, since such tests involve

new preparations, a regressive sequence of preparations and tests results, which, for practical purposes, must be truncated. Moreover, experimenters have to ensure that the specifications are maintained while the experiment is carried out, or allow for any changes that may occur. An experimental apparatus that gives correct results is considered good, but there appear to be no formal criteria that can decide the issue of whether or not an experimental apparatus is working properly. An appeal to tacit knowledge is paramount.

Calibration

This is the process of determining the correct functioning of an apparatus. The essence of calibration is the comparison between a known magnitude, serving as a standard, and a measurement result of the apparatus. Such a ground is important for developing convincing arguments regarding the trustworthiness of experimental results. For example, it can help truncate the experimental regress.

Continuity

Conceptual and theory change does not bring about changes in the data obtained from experimental setups, apparatus, and measuring devices. Progress in theory and in experimental technique is not necessarily synchronized. The functioning of hardware is robust with respect to changes in concepts and theories, so that data remain intact across such changes. This fact amounts to continuity; the persistence of experimental results provides continuity across conceptual and theoretical changes. The slogan "Experiment has a life of its own" captures this continuity. Moreover, a setup may develop its own "career," detached from the theory that brought it about, thus becoming itself an object of research.

Experimental Errors

Experimental data are notoriously corrupted by various kinds of error introduced at all phases of the experiment. The standard practice is to appeal to the dichotomy of systematic and random error, but, from a philosophical perspective, this is not helpful. The dichotomy does not focus on the source of the error; rather, a mathematical criterion is applied. The criterion distinguishes between statistical analysis and nonstatistical methods, and it does not reflect the complexity of experiment, an ensemble of theories, materials, instruments, measuring devices, and, of course, experimenters. Sources of error are shaped by the different phases of the experimental procedure. Moreover, experimental results may be proved wrong, but even in those cases where everything turns out right, numerous pitfalls and confusions have to be overcome.

Experimenters are dealing not only with errors but also with misguided conceptions, dead ends, and reorientations. Still, the scientific method requires the use of data to assess the values of theoretical quantities. How is this done? One idea is to use error statistics in order to obtain a philosophical understanding of reasoning in science. Two modes are recognized, namely, the evidential-relationship approach and the "severe-testing" approach. The former is Bayesian, whereas the latter is a development of the classical Neyman-Pearson statistics. Such approaches probe experiment with error by characterizing the experimental (or testing) process itself as expressing how reliably the process discriminates between alternative hypotheses.

Constructivism Versus Rationalism

Despite the complexity of the experimental scheme and the many possible sources of error, there appears to be an understanding that consensus can be reached on experimental results. For the constructivist, this is a matter of contingency, the consequence of social brokering among individuals and powerful communities of researchers. By contrast, the rationalist holds that consensus can be reached because experimental results are trustworthy; they are the consequence of convincing arguments that reflect nature. In short, general metaphysical issues can be brought to bear on experiment, thereby creating a new context for old philosophical issues and extending traditional discussions of realism and objectivity to the empirical base of science. This becomes paramount when large research groups experiment with enormous systems on an industrial scale.

Experimental Systems

Many modern experimental procedures, especially in biology and the social sciences, no longer consist of single, well-defined experiments. Experimenters who construct systems of

experiments deal with systems that are not well-defined and do not provide clear-cut answers. As the working unit of experimental science at the cutting edge of research, experimental systems are designed to allow for surprising answers to questions that experimenters are not yet able to articulate. The experimental system is essentially hybrid; it is an assemblage of elements such as research objects, theories, technical arrangements, instruments, as well as disciplinary, institutional, social, and cultural apparatus. Such systems are open and, by design, have no clear boundaries. They thus pose new philosophical difficulties.

Giora Hon

See also Bayesianism, Recent Uses of; Experiments in Social Science; Falsifiability; Logical Positivism/Logical Empiricism; Observation and Theory-Ladenness; Scientific Method; Social Constructivism; Tacit Knowledge; Thought Experiments

Further Readings

Buchwald, J. Z., & Franklin, A. (Eds.). (2005). *Wrong for the right reasons*. Dordrecht, Netherlands: Springer.

Franklin, A. (1986). *The neglect of experiment*. Cambridge, England: Cambridge University Press.

Franklin, A. (1990). *Experiment, right or wrong*. Cambridge, England: Cambridge University Press.

Galavotti, M. C. (Ed.). (2003). *Observation and experiment in the natural and social sciences*. Dordrecht, Netherlands: Kluwer Academic.

Galison, P. (1987). *How experiments end*. Chicago, IL: University of Chicago Press.

Gooding, D., Pinch, T., & Schaffer, S. (Eds.). (1989). *The uses of experiment*. Cambridge, England: Cambridge University Press.

Hacking, I. (1983). *Representing and intervening*. Cambridge, England: Cambridge University Press.

Heidelberger, M., & Steinle, F. (Eds.). (1998). *Experimental essays: Versuche zum experiment* (Interdisciplinary Studies/ZIF). Baden-Baden, Germany: Nomos.

Hon, G., Schickore, J., & Steinle, F. (Eds.). (2009). *Going amiss in experimental research*. Dordrecht, Netherlands: Springer.

Mayo, D. (1996). *Error and the growth of experimental knowledge*. Chicago, IL: University of Chicago Press.

Radder, H. (Ed.). (2003). *The philosophy of scientific experimentation*. Pittsburgh, PA: University of Pittsburgh Press.

Rheinberger, H. J. (1997). *Toward a history of epistemic things: Synthesizing proteins in the test tube*. Stanford, CA: Stanford University Press.

Shapin, S., & Schaffer, S. (1985). *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton, NJ: Princeton University Press.

Experimental Philosophy

As with most topics in philosophy, there is no consensus on what *experimental philosophy* is. Most broadly, experimental philosophy involves using scientific methods to collect empirical data for the purpose of casting light on philosophical issues. Such a definition threatens to be too broad, however; for instance, taking the nature of matter to be a philosophical issue, research at the Large Hadron Collider would count as experimental philosophy.

Others have suggested narrower definitions, characterizing experimental philosophy in terms of the use of scientific methods to investigate intuitions. This threatens to be too narrow, however, as it would exclude work such as Eric Schwitzgebel's comparison of the rates of theft of ethics books with the rates of theft for similar volumes from other areas of philosophy, for the purpose of finding out whether philosophical training in ethics promotes moral behavior.

While restricting experimental philosophy to the study of intuitions is too narrow, this nonetheless covers most of the research in this area. Focusing on this research, this entry begins by discussing some of the methods that have been used by experimental philosophers. It will then distinguish between three types of goals that have guided experimental philosophers, illustrating these goals with some examples. Experimental philosophy provides the ground on which certain social-scientific disciplines and empirical approaches are used to draw conclusions on philosophical issues. It is sometimes said that experimental philosophy challenges "armchair philosophy."

Methods

In a typical study in experimental philosophy, the researcher uses the methods of experimental psychology to study the intuitions of some group of people—most often the intuitions of people without

training in philosophy (often called "the folk"). This is usually done by constructing one or more vignettes, or hypothetical stories, that are of philosophical interest. The vignette(s) are then presented to participants, who are asked to answer one or more questions related to the vignette, often on a 5- or 7-point scale. The researcher then analyzes the results, considering what the participants' responses reveal about the intuitions at issue.

While this describes the majority of the studies conducted so far, other approaches have also been employed. For example, Adam Arico, Brian Fiala, Robert Goldberg, and Shaun Nichols have measured reaction times to investigate the low-level cues involved in mental state attribution. Adam Feltz and Edward Cokely have used a personality inventory to investigate the influence of individual personality differences on intuitions. And Jonathan Livengood, Justin Sytsma, Adam Feltz, Richard Scheines, and Edouard Machery relied on a social-psychological questionnaire—Shane Frederick's Cognitive Reflection Test—to highlight an aspect of the philosophical temperament: Philosophers tend to be more critical of their spontaneous, "gut" intuitions than non-philosophers (even when one controls for the level of education).

Furthermore, there are a number of psychologists conducting research that can be readily classified as work in experimental philosophy. These researchers often employ methods that go beyond the use of vignettes to solicit intuitions. For example, Joshua Greene and colleagues have used *brain imagery* in investigating the processes generating moral judgments.

Goals

It is common for philosophers to turn their analytic gaze on their own discipline, and experimental philosophers are no exception: There has been a good deal of debate concerning how best to classify various projects in experimental philosophy. The most prominent distinction is based on experimental philosophers' attitudes toward the use of intuitions as evidence in philosophy. Advocates of the positive program support this use of intuitions but hold that empirical investigation is often needed to arrive at an adequate understanding of what those intuitions are. In contrast, advocates of the negative program are skeptical of the use of intuitions as evidence (even when supplemented by empirical investigations), and their empirical work is intended to substantiate their doubts regarding this practice. Furthermore, some work in experimental philosophy is not motivated by the question of whether intuitions can serve as evidence, but rather by an interest in people's intuitions themselves and in what they reveal about how people think about topics such as free will, the mind, or consciousness. We will say that such work belongs to the neutral program.

While it is often relatively easy to identify the goals of experimental philosophers, it is sometimes difficult to classify their work as belonging to one program or another. To illustrate, consider Sytsma and Machery's work investigating how people classify different types of mental states. This research examined the philosophical claim that phenomenal consciousness is an obvious aspect of our mental lives. If this claim is correct, then non-philosophers should tend to classify mental states in the same way that philosophers do. Their experimental work suggested that this is not the case, however; unlike philosophers, non-philosophers treated two prototypical examples of phenomenally conscious mental states (seeing red and feeling pain) differently, suggesting that they did not view them as being phenomenal. Rather, non-philosophers seem to categorize mental states into different groups depending on the extent to which they have an associated hedonic value (i.e., related to pleasure).

How should this work be classified? It might be thought of as an example of the positive program insofar as it investigates people's intuitions and calls on those results to make a philosophical argument. Alternatively, it might be construed as an example of the negative program: Taking the philosophers' claim to rest on their intuitions about different mental states, this work challenges the reliability of that supposed evidence. Finally, having found evidence that non-philosophers do not classify mental states in the same way that philosophers do, Sytsma and Machery went on to explore how non-philosophers classify mental states, taking this to be an interesting question in its own right.

Despite this shortcoming, the distinction between the positive, negative, and neutral programs does a good job of capturing some of the most prominent projects in the literature. This is perhaps most clear with regard to the negative program. For example, Edouard Machery, Ron Mallon, Shaun Nichols,

and Stephen Stich presented evidence that intuitions about the reference of proper names vary across cultures. They then argued that, lacking a principled reason to favor the intuitions of one group over another, the evidential value of such intuitions is called into question. Similarly, Stacey Swain, Joshua Alexander, and Jonathan Weinberg have provided evidence that people's disposition to ascribe knowledge sometimes varies depending on a seemingly irrelevant factor—whether or not the situation is contrasted with a clear case of knowledge. They then argued that this raises doubts about the evidential value of intuitions about knowledge.

Other prominent work in experimental philosophy is well classified in terms of the positive program. Consider the work of Eddy Nahmias, Steve Morris, Thomas Nadelhoffer, and Jason Turner on incompatibilism—the claim that free will and determinism are incompatible. Incompatibilist philosophers often assert that it is counterintuitive to hold that free will and determinism are compatible, concluding from this that in the absence of a decisive argument to the contrary, incompatibilism should be seen as the default position. Nahmias and colleagues undermined this argument by providing evidence that ordinary people take free will to be compatible with determinism.

Turning finally to the neutral program, Joshua Knobe found that when judging whether someone intentionally brought about a foreseen side effect, people are more likely to judge that a harmful side effect was brought about intentionally than a helpful side effect. On the basis of these and similar findings, he concluded that moral considerations play a significant role in folk psychology.

Justin Sytsma and Edouard Machery

See also Epistemology; Folk Psychology; Moral Concepts; Cognitivism; Naturalized Epistemology; Thought Experiments

Further Readings

Knobe, J. (2006). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical Studies, 130,* 203–231.

Knobe, J., & Nichols, S. (2008). An experimental philosophy manifesto. In J. Knobe & S. Nichols (Eds.), *Experimental philosophy* (pp. 3–14). New York, NY: Oxford University Press.

Livengood, J., Sytsma, J. M., Feltz, A., Scheines, R., & Machery, E. (2010). Philosophical temperament. *Philosophical Psychology, 23,* 313–330.

Machery, E., Mallon, R., Nichols, S., & Stich, S. P. (2004). Semantics, cross-cultural style. *Cognition, 92,* B1–B12.

Nadelhoffer, T., & Nahmias, E. (2007). The past and future of experimental philosophy. *Philosophical Explorations, 10,* 123–149.

Nahmias, E., Morris, S. G., Nadelhoffer, T., & Turner, J. (2006). Is incompatibilism intuitive? *Philosophy and Phenomenological Research, 73,* 28–53.

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs, 41,* 663–685.

Schwitzgebel, E. (2009). Do ethicists steal more books? *Philosophical Psychology, 22,* 711–725.

Swain, S., Alexander, J., & Weinberg, J. M. (2008). The instability of philosophical intuitions: Running hot and cold on Truetemp. *Philosophy and Phenomenological Research, 76,* 138–155.

Sytsma, J. M., & Machery, E. (2010). Two conceptions of subjective experience. *Philosophical Studies, 151,* 299–327.

Experimenting Society, The

The idea of an *experimenting society* holds particular significance for philosophers and laypersons who seek greater understanding of the relationships between the methodology of the social sciences and public policy formation. The idea of social experimentation, while known to 19th-century thinkers such as August Comte and J. S. Mill, gained wide acceptance only in the era of the contemporary welfare state, when governmental programs were seen as appropriate means to solve problems of health, labor, education, and welfare. The idea of an entire experimenting society, however, originated in the writings of Donald T. Campbell (1916–1996), an applied social scientist and philosopher of science who during his remarkable 50-year career held regular and chaired professorships at the Ohio State, Northwestern, Syracuse, and Lehigh universities.

The Experimenting Society: A Methodological Utopia

In a widely circulated 1971 working paper titled "Methods for the Experimenting Society," the broad contours of the experimenting society were developed

as part of what Campbell viewed as an exercise in speculative utopian thought. Social scientists and policymakers, he argued, should attempt to find solutions to social problems through an open and self-critical process of social experimentation. Drawing on Karl Popper's *The Open Society and Its Enemies,* Campbell urged social scientists and policymakers to become a disputatious community of truth seekers.

Social scientists and policymakers would be servants of the experimenting society. They would justify their work on the basis of the seriousness of the problems facing societies, not on the basis of the supposed certainty of solutions to these problems put forth by social scientists, politicians, and "trapped administrators." There would be safeguards against the danger that social scientists become members of a self-serving elite, an elite that deliberately employs complicated statistical procedures and esoteric concepts and theories to render its ideas immune from criticism. Equally dangerous, however, is the false belief that social science advisors are inherently more knowledgeable than policymakers. In the experimenting society, policies would no longer be recommended as if they were certain to be successful, because the idea that one already knows precludes discovering through policy experimentation whether policies work or not.

Origins of the Experimenting Society

Contrary to conventional wisdom, the validity and practical relevance of social experimentation do not originate in the methods of the classical laboratory experiment or in the elegant field experiments conducted in the early 20th century to develop new statistical algorithms. This mistaken view is based on a failure to recognize that the experimenting society is not a mere extension of the randomized field experiment, a methodology developed for different purposes by Jerzy Neyman (1894–1981), a mathematical statistician, and Sir Ronald Fisher (1890–1962), a plant geneticist who, like Neyman, was a mathematical statistician.

On the contrary, Campbell based his work on early field experiments in education, including W. A. McCall's (1923) *How to Experiment in Education,* which presented views on randomization and research design that anticipated the work of Neyman, Fisher, and Gossett (aka "Student"). Their interests lay in theoretical subjects of probability and tests of statistical significance and not in improving schools, governments, or economies. The work of McCall was valued because of its practical contributions to knowledge about educational practice, not because of mathematical and statistical innovations such as random sampling, degrees of freedom, and the *t* and *F* distributions discovered by Gossett, Neyman, and Fisher.

The "True Experiment" as Ideal Type

The normative ideal of the experimenting society is the so-called true experiment—namely, an experiment in which an active intervention is randomly applied to randomly selected treatment and control groups, with the aim of establishing the causal relevance of the intervention (e.g., a new teaching method). The true experiment is a normative ideal by virtue of its specification of methodologically optimal procedures for testing the validity of causal inferences, inferences that relate one or more treatments, the presumed causes, to one or more treatment outcomes, the presumed effects.

The operational ideal of the experimenting society, however, is the *quasi-experiment,* namely, an experiment conducted outside the laboratory in a field setting where the random assignment of policies to randomly selected persons, groups, or contexts is impractical, illegal, or unethical. Although quasi-experiments may involve random selection of treatment and control groups (but not random assignment of a treatment to these groups), they are designed in and for complex field settings where manifold contingencies lie beyond the control of experimenters.

In contrast to the essentialist notion of causality, which contends that causal inferences must be based on joint necessary and sufficient conditions, the experimenting society proceeds from a recognition that a variety of uncontrollable contingencies are responsible for the complex causal texture of field settings, in contrast to the relatively simple causal texture of laboratory settings. This notion of *representative design* is important for understanding the differences between the simple problems addressed by true experimental design, and the complex problems addressed by quasi-experimental design.

The Theory of Representative Design

The theory of representative design, developed by Edward Tolman and Egon Brunswik, Campbell's

mentors at Berkeley, contends that optimally valid experiments should be carried out in the typical ecology in which interventions are expected to produce effects. The theory of representative design implies that it is not possible to control all or most contingencies, for example, through a standard pretest–posttest design with random assignment to treatment and control groups. To be sure, randomized clinical trials and other randomized experiments are carried out in field settings—for example, randomized clinical trials in biomedical research or agronomy. Nevertheless, the experimenters must contend with unmanageable contingencies by identifying and ruling out the influence of factors that lie outside their direct control. These effects are known as threats to validity, or plausible rival hypotheses, which typically combine with a policy intervention to render causal inferences equivocal.

The genius of Campbell and his collaborators was the insight that these unmanageable contingencies could be placed in a multifold classification scheme that contains a practically comprehensive set of more than 35 specific threats to the internal, external, construct, statistical, and context validity of causal inferences. The principal source today is Shadish, Cook, and Campbell's *Experimental and Quasi-Experimental Designs for Generalized Causal Inference,* a book that followed earlier seminal contributions in the 1960s and 1970s by Campbell and Stanley in *Experimental and Quasi-Experimental Designs for Research* and by Cook and Campbell in *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Together, these works are among the most widely cited sources on the methodology of field experiments in the social sciences.

Beyond Logical Positivism

At first glance, the methodology of quasi-experimental design in field settings may appear simply to be a variant of econometrics and other quantitative social science disciplines. However, quasi-experimental design, while employing multivariate statistical procedures to test and rule out rival hypotheses, is grounded in a post-positivist epistemology, methodology, and sociology of the social sciences. The post-positivist philosophy of the experimenting society stands in sharp contrast to the still thriving logical positivist sentiments of many social scientists, notwithstanding the abandonment of logical positivism by most philosophers of science by the mid-1950s.

A cardinal post-positivist tenet is the rejection of certainty as an attainable goal and its replacement by inductive plausibility as a defining characteristic of policy-relevant knowledge. For Campbell, the experience of "reality" is indirect and distal, mediated by a host of factors—psychological, social, political, economic, cultural—that render unattainable any notion of a complete correspondence of hypothesis and observation, word and object. In the experimenting society, we learn vicariously, by means of independently imperfect theories, methods, measures, and observers. The discovery of plausibly true beliefs—as distinguished from the justified true beliefs required by logical positivism—is an unending quest for approximate and socially contingent truths.

Coherence Theory of Truth

Policy-relevant knowledge demands a commitment to a particular philosophical theory of truth, the coherence theory of truth, which is distinct from the correspondence theory. The coherence theory asserts that truth arises from a comparison of two or more empirically grounded beliefs, in contrast to the correspondence theory, which asserts that truth arises from the agreement of beliefs with an erstwhile objective reality. In this context, social scientists are ethically bound to investigate the degree to which their empirically grounded beliefs cohere with those of other members of the experimenting society. The coherence theory rejects the logical-positivist claim that single operational measures define scientific concepts (mono-operationism). Instead, approximations to truth are achieved through multiple operational definitions (multiple operationism) and a strategy of multiple triangulation. In testing multiple theories, methods, observers, and observations, social scientists employ eliminative rather than enumerative induction. It is the falsifiability of knowledge claims, rather than their confirmability, that governs the process of testing and eliminating rival hypotheses, or threats to validity. These threats arise from the many unmanageable contingencies facing the experimenter.

The structures and processes of the experimenting society are closely related to the coherence theory of truth. The ideas of social scientists, policymakers, and citizens are brought together and examined

for their coherence, or lack thereof, in providing causally relevant explanations of why policies succeed or fail. Here, we see that experimentation is not the sole basis of what has come to be known as evidence-based policy; experimentation is not necessarily contrary to traditional wisdom based on the evolving experience of policymakers. The experimenting society is grounded in *evolutionary epistemology,* with experimentation best seen as a refining process superimposed upon cumulations of wise practice.

William N. Dunn

See also Evidence-Based Policy; Experiments in Social Science; Falsifiability; Mathematical Models, Use in the Social Sciences; Models, Induction and Confirmation; Models in Social Science; Popper's Philosophy of Science; Truth, Philosophical Theories of

Further Readings

The premier source on evidence-based policy is the Campbell Collaboration: http://www.campbellcollaboration.org

Campbell, D. T. (1989). *Epistemology and methodology for social science: Collected papers* (E. S. Overman, Ed.). Chicago, IL: University of Chicago Press.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.

Dunn, W. N. (1997). Pragmatic eliminative induction: Proximal range and context validation in applied social experimentation. *Philosophica, 60,* 75–112. (Special issue dedicated to the memory of Donald T. Campbell)

Dunn, W. N. (1999). *The experimenting society: Essays in honor of Donald T. Campbell*. New Brunswick, NJ: Transaction.

Shadish, W., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Experiments in Social Science

This entry reviews the kind of experimentation used in social science, delineates its essential aspects, and explains its role as well as the kind of validity appropriate in this case.

Controlled experiments in laboratory conditions are used in various branches of the social sciences. Whereas in some disciplines, like economics, experimentation is now part of the mainstream, in others, like political science, it is gaining a foothold with difficulty. Meanwhile, laboratory research is still considered marginal and is used only rarely by sociologists and anthropologists. Despite its controversial status, however, experimentation has always exerted an influence on those disciplines, like psychology, that use it regularly for the generation of scientific knowledge.

Measurement and Causal Inference

Experiments are designed for various purposes, the most important ones being *measurement* and *causal inference*.

When experiments are used as measurement devices, the goal is to detect and quantify a variable of theoretical interest. For example, one may want to measure the degree of cooperativeness of individuals belonging to a certain population, or their prosocial preferences in a controlled setting. *Replicability* and *robustness* are important virtues of experimental measures: Different research teams following identical protocols should generate statistically identical measures (replicability), and it should be possible to obtain the same measures using different measurement techniques (robustness).

Measurement in social science is rarely aimed at the determination of theoretical constants that hold universally but rather at the discovery and quantitative testing of *causal hypotheses*. In a competently performed experiment, each treatment or condition is designed so as to introduce variation in one (and only one) potential causal factor. Controlled experimentation thus allows underlying causal relations to become manifest at the level of empirical regularities. In a competently performed experiment, single causal connections can be "read off" directly from statistical associations.

Causal discovery requires variation—but not too much variation—and of the right kind. In general, one wants variation in one factor while keeping all the other putative causes fixed "in the background." This logic is exemplified in the *model of the perfectly controlled experiment* (see Table 1).

Table 1 The Perfectly Controlled Experimental Design

	Treatment (Putative Cause)	*Putative Effect*	*Other Factors (K)*
Experimental group	*X*	Y_1	Constant
Control group	—	Y_2	Constant

Source: Author.

The *Ki* are background factors, or other causes that are kept fixed across the experimental conditions. The conditions must differ with respect to just one factor (*X*, the treatment), so that any significant difference in the observed values of *Y* (Y_1–Y_2) can be attributed to the presence (or absence) of *X*. A good experimenter thus is able to discover *why* one kind of event is associated regularly with another kind of event—and not just that it does. In the model of the perfectly controlled experiment, one does not simply observe that "if *X* happens then *Y* happens" or even that "*X* if and only if *Y*." Both conditionals are material implications, and their truth conditions depend on what happens to be the case, regardless of the reasons *why* it is so. In science in contrast—and especially in those disciplines that regularly inform policy making, like the social sciences—one is also interested in "what would be the case if" such and such a variable was manipulated. Scientific intervention and policy making must rely on *counterfactual reasoning*. A great advantage of experimentation is that it allows checking what would happen if *X* were *not* the case, while keeping all the other relevant conditions fixed.

This marks an important *difference* between the experimental method and traditional statistical inferences from field data. Using statistical techniques, one can establish the strength of various correlations between economic variables. But except in some special or opportune conditions, the spontaneous variations found in the data do not warrant the drawing of specific causal inferences. This does not mean that total experimental control is always achieved in the laboratory. The perfectly controlled experiment is an idealization, and in reality there are always going to be uncontrolled background factors, errors of measurement, and so forth. To neutralize these imperfections, experimenters use various techniques, like, for example, *randomization*. In a randomized experiment, subjects are assigned to the various experimental conditions by a chance device, so that in the long run the potential errors and deviations are evenly distributed across them.

Internal and External Validity

Scientists are aware that the successful discovery and testing of causal claims does not automatically ensure that these claims can be generalized to nonexperimental circumstances. For this reason, distinguish between the *internal* and the *external validity* of experimental results. Problems of internal validity have to do with the drawing of inferences from experimental data to causal mechanisms in a given laboratory setup. External validity problems instead have to do with the drawing of inferences from experimental data to what happens in other (typically, nonlaboratory) situations of interest. External validity inferences rely on the effective combination of field and experimental data. Recent approaches to external validity include *analogical reasoning* (propounded by Guala) and *comparative process tracing* (propounded by Steel).

Francesco Guala

See also Causation in the Social Sciences; Cooperation/Coordination; Econometrics: Methodological Issues; Experiment, Philosophy of; Scientific Method; Thought Experiments

Further Readings

Guala, F. (2005). *The methodology of experimental economics*. New York, NY: Cambridge University Press.

Guala, F. (2008). Paradigmatic experiments: The ultimatum game from testing to measurement device. *Philosophy of Science, 75,* 658–669.

Kagel, J. H., & Roth, A. E. (Eds.). (1995). *The handbook of experimental economics*. Princeton, NJ: Princeton University Press.

Morton, R., & Williams, K. (2008). Experimentation in political science. In J. Box-Steffensmeier, H. Brady, & D. Collier (Eds.), *The Oxford handbook of political*

science (pp. 339–356). New York, NY: Oxford University Press.

Steel, D. (2007). *Across the boundaries: Extrapolation in biology and social science*. New York, NY: Oxford University Press.

Woodward, J. (2009). Experimental investigations of social preferences. In H. Kincaid & D. Ross (Eds.), *The Oxford handbook of the philosophy of economics* (pp. 189–222). New York, NY: Oxford University Press.

Experiments in the Social Sciences: Ethical Issues

Experiments, long the standard research setting in biology, chemistry, and physics, are increasingly important in the social sciences for testing derivations from theories. Experiments are research designs in which the investigator controls the level of one or more independent variables *before* measuring the dependent variable(s). The *control* and *time* elements distinguish experiments from other designs common in the social sciences, such as observation studies and ethnographies, surveys, content analyses, and historical descriptions.

Control, of course, introduces ethical issues, including responsibility of the experimenter. Experiments in the social sciences usually involve human participants, and that fact too entails ethical considerations. (Animal experimentation, not dealt with here, involves some comparable and some different ethical considerations.) The overriding ethical concern with human experimental participants is to protect their well-being throughout, even, where possible, improving it through new insights and experiences.

Professional and Governmental Guidelines

Professional societies, to which almost all faculty and organizational researchers belong, publish ethical guidelines for experiments involving humans. For instance, the American Psychological Association (APA) requires that psychologists "respect the dignity and worth of all people, and the rights of individuals to privacy, confidentiality, and self-determination," and the American Sociological Association (ASA) echoes that principle. APA, ASA, and other professional associations spell out the meanings of that principle in detail. Since 1979, U.S. governmental funding agencies, which support much of the social science research in the United States, have subscribed to a "Common Rule" for research protections. The European Parliament has passed some comparable regulatory requirements, though emphasizing pharmaceutical research.[1] While such guidelines go a long way toward ensuring ethical treatment, like all rules, they depend on interpretation and application by well-intentioned and knowledgeable researchers.

Issues in Designing Research With Humans

Designing experiments requires thorough planning, and it is helpful to remember the purpose of experimental designs: It is to compare outcomes (dependent variables—usually behavior and cognitive responses) among individuals in different groups that are alike except for the factors (the independent variables) under test. The requirement of similarity means that an experiment properly begins with the first contact—usually the recruiting of participants—and does not end until they have been escorted to the hallway and thanked after their participation. Experimenters should stick to scripts as much as possible. In particular, the recruiting process must be uniform; all potential participants should be solicited in the same ways, given the same information, offered the same incentives for participation, and so on.

Incentives

Participants may be solicited with a number of incentives, such as money, interest, and course credit. In the United States, introductory psychology courses routinely require students to sign up for research or other projects. Other social sciences and nonuniversity research organizations in the United States and most researchers in Europe are more likely to offer monetary payment. Naturally, institutional review boards (IRBs) look especially closely at proposed research with "vulnerable populations," including schoolchildren and prisoners, who may not be completely free to choose whether to participate. Studies have not found consistent effects of the main distinction —payment or course credit as incentives—so long as participants are treated considerately and respectfully, as, of course, they always should be.

Besides moral concerns, if participants feel that they have been mistreated or coerced, they can always behave in a bizarre fashion and produce misleading data. On the positive side, it seems that most participants volunteer out of interest and to

learn, rather than for money or course credit. When part of the incentive is entry in a drawing for money at the conclusion of the experimental runs, it often happens that not all of those who are notified turn up to claim their winnings.

General Design Issues

Along with scientific issues of design and operation, there should always be a concern for respectful, considerate treatment of the people who volunteer to participate in research. This extends even to a seemingly small matter such as word choice. People are "participants," not "subjects"; they will participate in a "study," not an "experiment," which has unwanted connotations for some; it is more respectful to ask how someone "responded" to something rather than the more passive "reacted." Researchers should keep in mind common sources of discomfort, such as potential public embarrassment or violations of privacy, and do whatever they can to assure participants that the research will not bring them those unwanted consequences.

Even with such attention to detail, many experimental designs necessarily include stressful aspects. Some ask participants to work on tasks at which they may do poorly; others ask them to make decisions pitting their own benefits against other people's. It is the researcher's responsibility to anticipate, reduce, and afterward to overcome any stressful elements of the experience. Participants deserve a complete post-session interview and explanation. The researcher must explain all aspects of the experiment—what the arrangements were and why, what the research questions were, and how the participant helped. The explanation must be in terms accessible to a normal member of the population, not in technical terms. Each person who participated in an experiment must feel that he or she has a good understanding of what he or she did, why the experiment was designed as it was, how the information is useful, how he or she contributed to the work, and what he or she gained from the experience.

Well-planned individual interviews are best for explaining an experiment, although sometimes that is not practical—for instance, when the number of participants in each group exceeds the number of researchers. In such cases, a group explanation, oral or written, may be used. Pretesting, in which individual interviews are conducted, is especially important here. In every case, whether individual or group explanations, each participant should be made to feel that he or she can ask for any clarification. Also, participants should be provided a contact point, such as an e-mail address, in case they have any questions later.

Deception

Some social science experiments are deceptive (e.g., the Milgram experiment—see below, also the Asch experiment); the designs do not always tell participants everything about the situation prior to their participation, and sometimes the design includes giving false information. There are many reasons for designs using deception. Sometimes the reason is to avoid self-presentation concerns affecting behavior: Studies of racial and gender bias are particularly sensitive to that. Or a design may test a theoretical derivation that occurs only rarely in natural settings, such as how often people will conform to others who they know for sure are saying something untrue. Sometimes the reason is efficiency: A theory under test might predict how likely interactants are to accept influence when someone disagrees with them, and a research design might therefore tell interactants that they are in disagreement much more often than they actually are, to avoid wasting everyone's time with instances where they do not disagree.

For those and other reasons, deception is a feature of experiments in most of the social sciences. Ethical codes and government regulations permit deception if it does not unduly stress or otherwise endanger participants, if there is no nondeceptive way to study the phenomena, and if the benefits to participants and to society outweigh possible negative effects on participants. Furthermore, the deceptive aspects must be fully explained as soon as possible after the data are collected. Deceptive research designs receive special scrutiny from IRBs, for obvious reasons.

Contamination

Contamination means something simple in reference to experimental research: talking about it in a way that hurts new data collection. Suppose that a research design involves deception, and suppose further that someone who has participated tells her friends about it. If they participate in the study later, they won't believe some elements, and so data from their sessions may be useless. Worse, even if they later participate in a nondeceptive design, they may

be suspicious, and that could affect their responses. The population would be contaminated.

Many economists, including economists who conduct experimental research, believe that deception should never be a part of experiments for that reason. They argue that any deception will contaminate a population, such as the student body of an entire university, and so make results of further experimental research invalid. Of course, it would not matter whether the deception causing contamination was done by an economist or, say, a psychologist, so this perspective leads to a view that nobody should be allowed to do research including deception.

This issue is far from settled. However, two points may be helpful.

First, the objection presumes that contamination will follow deception. Whether or not that happens depends on whether participants feel that they have been mistreated. If they do feel mistreated, it is easy to shut down a project by talking about it. On the other hand, if the researcher has treated participants well, they are likely to want to help with the research and keep confidence about it if they are asked to do so. The economists' objection is on empirical grounds, not ethical, and we do have evidence. Psychologists and sociologists have been conducting experiments, some of them including deception, for more than half a century, and reports of contamination hindering the research are rare or nonexistent.

Second, the ethical issue here is not whether to use deception. Economists' objections are on practical grounds, contamination, not on whether deception is ethically acceptable. The ethical issue is how much right one group of scholars—economists—has to restrict the activities of others—psychologists, sociologists, political scientists, and the rest. On one side, if, say, sociologists contaminated a pool, that would adversely affect economics researchers, and so they ought to have a say. On the other side, the issue of one scholarly group telling another how to conduct their lives may be troublesome. This debate seems unlikely to be resolved soon. One positive outcome is to underline the importance of respectful treatment of participants throughout, including full explanations afterward.

"Risk" and Psychological Stress

Historically, U.S. government rules for the protection of human subjects arose in response to some egregious misuse of humans in medical, not social, research. Among the protections was a requirement for *informed consent.* A person must know about a proposed treatment, any alternatives, and possible risks and benefits before agreeing to participate. While informed consent is essential for medical studies, it is not a full solution to ethical issues in social research.

The possible risks of participation in social research usually are social-psychological rather than physical: anxiety, discomfort, stress, worry about self-presentation, and so on. As noted above, researchers have the duty to anticipate and minimize these stresses when designing experiments. Why is informed consent not a full solution? Because most of us are not very good at anticipating how we will respond to new situations, or even whether a particular situation will be stressful for us.

In a notorious experiment conducted by the psychologist Stanley Milgram some decades ago, participants were told that they must administer electric shocks to another participant to punish him when he gave wrong answers on a test. The shocks were not real, and that fact was explained to participants afterward, but during the experiment the simulated pain of the person supposedly being shocked was vivid. If we ask college students now how that would affect them once they learned the shocks were not real, they say that they might be worried during the session but would feel fine once they learned the shocks were not real. Yet in the actual case, some participants reported sleep problems, even nightmares, weeks afterward.

A similar effect has been reported from much less dramatic experiments where participants played a game in which either they could win money by contributing to a group pool that would be split or they could win even more by failing to contribute while everyone else contributed. (Technically, this is a "mixed-motive deficient-equilibrium" game. This research did not involve deception.) Some participants who won by failing to contribute called the researchers days or weeks afterward, again reporting significant bad feelings and wanting to return the money they had won. The point is the same: We are not good at estimating how stressful strange experiences will be for us.

Therefore, it is the responsibility of researchers to anticipate and to minimize or eliminate sources of stress. Informed consent of participants, while important, is not enough. Researchers, social scientists by training and experience, ought to know better than untrained college students or other citizens

which aspects of a particular experiment might be stressful and how to deal with those. Again, researchers have an ethical responsibility to know what to do to protect participants and to implement that knowledge in research design and operations.

Ethical Experiments

The experimental method is not inherently ethical or otherwise, although particular experiments might be designed and conducted ethically or otherwise. Every step of the research process requires planning and monitoring; there is no once-and-for-all way to promote ethics in experimental research. Experimental researchers enter an ethical contract with humans who volunteer, even for incentives, to participate in the enterprise. To be worthy of that trust, researchers must use knowledge, sensitivity, and respect for their participants and for other scholars. In other words, researchers must fully acknowledge their humanity.

Murray Webster

See also Ethical Impact of Genetic Research; Experimenting Society, The; Experiments in Social Science; Policy Applications of the Social Sciences

Note

1. Guidelines of several American government agencies, professional societies, and the European Union on ethical requirements can be obtained from the following websites:
   http://www.nsf.gov/bfa/dias/policy/hsfaqs.jsp#relation
   http://www.nsf.gov/oig/dearcolleague.pdf
   http://www.apa.org/ethics/code/index.aspx
   http://www.asanet.org/about/ethics.cfm
   http://www.apsanet.org/content_9350.cfm
   http://ec.europa.eu/research/biosociety/bioethics/bioethics_ethics_en.htm

Further Readings

Asch, S. (1951). Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (Ed.), *Groups, leadership and men.* Pittsburgh, PA: Carnegie Press.

Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science, 326,* 535–538.

Joel, I. (Ed.). (2011). The human sciences and Cold War America [Special issue]. *Journal of the History of the Behavioral Sciences, 47*(3).

Lemov, R. (2006). *World as laboratory: Experiments with mice, mazes and men.* New York, NY: Farrar, Straus & Giroux/Hill & Wang.

Sell, J., Hertwig, R., Ortmann, A., Cook, K. S., & Yamagishi, T. (2008). Debate on deception in experiments. *Social Psychology Quarterly, 71,* 213–221.

Webster, M., & Sell, J. (2007). *Laboratory experiments in the social sciences*. New York, NY: Elsevier.

Wilson, T. D., Aronson, E., & Carlsmith, K. (2010). The art of laboratory experimentation. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, pp. 51–81). New York, NY: Wiley.

Explanation, Theories of

The concept of *explanation* as a subject for philosophical study, distinct from that of *knowledge* or *science,* came into its own in the 20th century and was fundamentally shaped by the initial contributions of Carl Hempel and Paul Oppenheim, although it can be argued that many systematic thinkers, such as Aristotle, held views that deserve attention in connection with the subject.

A theory of (scientific) explanation must answer these questions: What constitutes an explanation in science; whether different scientific subdisciplines proffer different products by way of explanation; whether therefore there are different species under the genus of explanation, and if so how they are related; and also how the genus of explanations that qualify as scientific differ from anything else that are referred to as explanations outside such contexts.

This entry surveys briefly the top contenders by way of philosophical theories of scientific explanation and then examines some of the distinctive philosophical issues raised for the topic of explanation by the social sciences.

Some History of the Topic

Aristotle's theory of science was one and the same as his theory of knowledge, which he understood as inquiry for its own sake. He held that inquiry falls into natural categories that today we might refer to as disciplines, and his primary contrast was that

between *knowledge* (the product of inquiry for its own sake, needing to serve no further end) and *common sense* (which serves many useful ends). Aristotle maintained that standards of inquiry varied across disciplines: for example, that logic and mathematics had standards of proof that differ from standards of evidence in the study of nature or in politics (which for Aristotle was also a discipline). And he was clear that common sense is not a form of knowledge because common sense has no standards at all. It is not a discipline but simply the result of a kind of practical activity of human life and so is to be evaluated by its serviceability in those activities.

Aristotle's theory of science had sought to answer the two most important questions in the area of (scientific) explanation: (1) How does a (scientific) explanation of something differ from a mere description or explication (or anything else) of it? (2) What makes such an explanation actually scientific, rather than something else? Aristotle's answers are that (1) scientific inquiry results in an understanding of essences, first principles, and definitions of key (discipline-defining) concepts that can be organized in syllogistic form and that answer the question "Why?"and (2) scientific inquiry is governed by norms orienting it toward truth—norms that do not govern common sense.

However, until we are presented with a clear sense of what the truth-orienting norms of science are and how it is possible to fall afoul of them, any pursuit or community of seekers can claim that it is governed by truth-orienting standards—from phrenology to the various forms of divination—particularly if the so-called norms explicitly provide that the uninitiated have no access (legitimate or otherwise) to its true (and possibley also actively shrouded) practice. Until philosophical light illuminates the norms associated with science proper, anything practiced in the shadows could raise itself to an unwarranted height. This is the gap that Karl Popper sought to fill in the early part of the 20th century—and for which he has rightly been widely esteemed by the scientific community.

Popper made it his fundamental concern to demarcate between science, as such, and things of lesser dignity. To this end, he focused on the idea of scientific progress—that certain ideas could be advanced ahead of others that precede them in order of appearance before the tribunal of science. Like many of his contemporaries, Popper took note of the fact that observation alone could not advance a hypothesis (of scientific credentials) ahead of its competitors, since any given body of observations will be consistent with numerous and mutually incompatible bodies of theory. And he brilliantly maintained that the core of a philosophy of science would account for how an idea, however initially outrageous in its original context, can grow in scientific stature, to the point of eclipsing or even displacing older, better-positioned, and perhaps even better-regarded theories. This account would spell out the method of scientific comparison, of the testing of one eligible hypothesis against another. For the distinction of "scientific method," Popper proposed the method of *falsification*. According to this account, theory is scientific to the extent that it excludes or prohibits certain possibilities that are in principle observable—and is ipso facto falsifiable; and conversely, it is corroborated to the extent that it survives tests aimed at falsifying it. Popper's critics were later to argue that the method does little in the way of providing a theory of comparisons—it is merely a measure of the severest test of a hypothesis and not against other theories but against nature itself. And rarely are the conditions for such a severe test available, because hypotheses can be shielded from falsification in numerous ways (that we need not discuss here). The point to note here is simply that Popper's contributions occasioned the philosophical community to mark an important distinction vis-à-vis the functions of scientific theory: On the one hand, scientific hypotheses receive attention as objects of scrutiny—by being subjected to *confirmational testing* (and of course on this matter, Popper's proposal is that they be subjected to *falsification*); and once so scrutinized (and surviving scrutiny), scientific hypotheses might subsequently be called on to perform a range of other services to which the label of *explanation* might be appropriate. Not everything eligible for scrutiny (eligible for conformational testing) is called on to perform the task of explanation. Theories of explanation take up the subject of what Popper leaves largely unelucidated in Aristotle's theory of science.

The first—and still most influential—theory to tackle the question of explanation in the 20th century is quite closely aligned with the first of Aristotle's answers. The *deductive-nomological* (DN) model of explanation was put forward by

Hempel and Oppenheim. The DN model drew an enormous quantity of criticism—all testifying to the philosophical importance of getting the account of explanation right. Some critics (e.g., W. Salmon and H. Kyburg) charged that DN allows too much to count as scientific explanation; others (e.g., M. Scriven) claimed that it disallowed too much from qualifying as genuine scientific explanations. The criticism of DN sparked a large literature, indeed an entire branch or industry in philosophy, that continues to bear fruit to this day. Efforts to meet some of the criticism resulted in the *causal-mechanical* (CM) model of explanation and the *unificationist* model of explanation—both attracting criticism of their own.

The CM model, propounded by Wesley Salmon in 1984, reflects the influential ideas on causation first expressed in the work of Hans Reichenbach. There have been a number of different ways of filling out the details of a CM model, each with different strengths and weaknesses but all sharing the fundamental idea that explanation of something must cite its causes. The different articulations differ in how they conceptualize a cause but are united in the single thought that DN explanations fail precisely when they neglect to cite causes. Of course, the CM models also have their critics. One of the most influential has been Philip Kitcher, who advanced instead the conception of explanation as the drawing together, into a more unified account or formulation, of a range of different phenomena. This would show how a range of apparently different facts or events can be profoundly illuminating. And such demonstrations have doubtless played an important role in the history of science: Isaac Newton's theory of gravitation wrought a unification of terrestrial and celestial motion, and James Clerk Maxwell's electromagnetic theory wrought a unification of electricity and magnetism—and the theory of light. Such unifications are greatly important for the advancement of science, but one might wonder whether unification is always required for explanation.

Another prominent contender on the topic of explanation in science (the last we will examine here) takes this question seriously, answering that there is no one single requirement of explanation. The logic of explanation, on this account due to Bas van Fraasen, is the logic of question and answer—more specifically, the logic of answers to why-questions. And since the proper answer to a given why-question depends on the reasons for which the question was posed in the first place, the logic of answers to why-questions must do so as well: There are a number of different profiles of question-and-answer communications, each with its own standards and requirements. Requests for explanation thus fall under a logic of communication.

The Genus Conception

The sciences have proliferated since the beginning of the 20th century. And with a growing appreciation of this variety that is science, philosophy is itself maturing vis-à-vis the philosophical topic of scientific explanation. A recent move in this philosophical space has been to suggest that explanation might be a genus, with several species falling under it sharing nothing especially obvious in form or overt presentation. Furthermore, to identify and articulate the form or structure of each species of explanation, one must gain an appreciation of the producing scientific subdisciplines and their subject matters. With this and related ideas, reductionism as the dominant approach in the metaphysics of science is also waning. Rather than espousing the idea that the laws of nature are basically the laws of physics, and that everything else comes in some sense "for free" once these are in place, philosophers are now in large numbers endorsing the notion of *emergence,* according to which new features (and so also new laws of nature) appear at a variety of levels of organization of matter. And these laws are not in some sense already contained in the laws of physics. In other words, theories devised in the discipline of physics are emphatically not theories that can even remotely cover everything.

Issues Raised by the Social Sciences

The social sciences bring a host of topics to the table. We have already mentioned the larger issue of reductionism, which bears also on the social sciences: Are the phenomena that are salient at a scale of social organization simply the result of more elemental facts about the relevant organisms, their constituents, and the situation? Are they, in other words, a kind of cipher for phenomena that belong more properly to the sciences that deal with smaller-scale entities, or do they rather possess a dignity of their own, demanding explanations that draw on features that appear only at larger scales of analysis?

These are questions that can be asked of any scientific enterprise outside physics (or even inside it). But associated with the social sciences are issues that get entangled with these larger issues in special ways, a topic to which we will now turn.

Are Social Sciences Different From Natural Sciences?

Economists by and large take individual consumption or investment behavior as the basic datum of their discipline; and sociologists include other categories of individual human behavior as basic. More generally, the social sciences take human individuals as the "fundamental" units of study (the smallest unit of the world that generate relevant data for their disciplines), even if their particular disciplinary focus is primarily on institutions and large-scale properties of social organizations or communities—in much the same way that the fundamental units studied in chemistry are atoms, even if the focus in the discipline is primarily on chemical processes and their properties on a larger scale. This truth is a double-edged sword. On the one hand, it provides social scientists with tools that enhance insight into a given phenomenon under study (the scientists can utilize such insights as they might have into human behavior, perhaps from first-person experience). On the other, it suggests a divide—indeed a veritable chasm—between the natural and the social sciences.

A school of thought in the 20th century called *interpretivism,* brought to prominence in the writings of the British philosophers R. G. Collingwood and Peter Winch and tracing its origins to Wilhelm Dilthey (1833–1911), has taught that natural and social sciences are fundamentally different: The former deal with broadly causal processes, whereas the latter do not. The character of the phenomena with which social sciences deal is construed by thinkers in this school as importantly different because their springs or grounds are what we refer to in common parlance as "reasons." Reasons also give actions their *meaning*—something that is not local to the action itself. Reasons have their life in the practices of redescription, which allow us to grasp the point (the purpose) of an action within its setting of rules, practices, conventions, and expectations—in other words, to place it in a wider setting that affords evaluation as to its rationality or as to some other form of appropriateness. And this—as they insist—is a normative rather than a purely descriptive or explanatory setting. And so the relations between an action and its ground cannot, on this view, be purely descriptive, or purely causal. And since meanings cannot be left out of any complete account of human behavior, a study of meanings cannot be purely a naturalistic discipline. An explanation in social science must always draw on acts of interpretation performed by (human) scientists.

And this suggests a parallel between the social and the natural sciences, vis-à-vis the topic of reduction: If natural sciences are beholden in a fundamental way to physics as ontologically the most basic, social sciences are beholden in a fundamental way to human action as ontologically basic. The phenomena of social sciences might then be reducible to that level—but no further. Indeed they *must* be reducible to that level, as yet another set of teachings declared.

Max Weber (*Economy and Society* [1978]), founder of the dominant *verstehende* school of social science, addressed a fundamental question of methodology in sociology—the question of how to conceive of actions and the agents who author them. Regarding multi-individual collectivities and institutions, he wrote, "In sociological work these collectivities must be treated as solely the resultants and modes of organization of the particular acts of individual persons, since these alone can be treated as agents in a course of subjectively understandable action" (1978, p. 13). In other words, Weber proposed reserving the term *action* for human behavior that proceeds from the machinations of an individual human mind that conceptualizes (or, in the language of the school Weber would eventually establish, *understands*) the behavior in question in a particular way. This concept of action is, far and away, the most enduring legacy of the *verstehen* school of social science. It issues in what is now referred to as *methodological individualism,* according to which the individual is the one and only unit of agency, because (again according to the doctrine) the individual is the one and only unit in which meaning and understanding, as rational enterprises, are manifest. This doctrine now dominates the social sciences, having worked to displace methodologies sympathetic to classical sociology. Many proponents of the idea propose Decision Theory, under an individualistic reading, as the explanatory fundamental of all social sciences.

While the doctrine of methodological individualism is widely criticized today, it is also widely embraced (at least where it is proposed that individualistic Decision Theory must be the queen of the social sciences), in some cases by the same researchers. And this too cries out for philosophical attention.

Statistical Laws

Émile Durkheim (1858–1913) was a pioneer of the discipline of sociology in France. In 1897, he published a seminal work, *Suicide,* a demographic study of the differences in suicide rates among Catholic and Protestant populations in Europe. The data indicated lower suicide rates among Catholics. And these facts occasioned an ingenious application —perhaps even a test—of the concept most closely associated with Durkheim's name: the concept of *solidarity* (or *integration* as it is sometimes translated). Durkheim argued that the higher level of solidarity among Catholic populations tends to protect against causes of suicide.

Durkheim reasoned that in modern societies, by contrast with traditional ones, the highly complex division of labor results in "organic solidarity." This is a condition in which different specializations in employment and social roles create dependencies that tie people to one another. In less modern societies, which he referred to as "mechanical societies," held together by "mechanical solidarity," subsistence farmers live in communities that are self-sufficient and knit together by a common heritage. Mechanical solidarity thus comes from homogeneity, when people feel connected through similar work, educational and religious training, and lifestyle. Mechanical solidarity is thus wrought by a sameness of mind and thought. However, an increasing division of labor begets a more diverse and varied individual consciousness. "Individualism" emerges as distinct from collective or common consciousness. And the individual often finds himself or herself in conflict with such collective or common consciousness as there might sometimes seem to be. This proposition can thus lead to predictions, as much as it leads to explanations, of observed data.

But how, precisely, does a proposition (e.g., to the effect that higher levels of integration protect against suicide) explain an instance of suicide? When someone from a Catholic community, or a Protestant community, commits suicide, how does Durkheim's theory apply? Should Durkheim's theory be understood as explaining the suicide of a given Catholic or a given Protestant recorded in the data? How can it explain both? And what exactly can it mean to say that a certain suicide was due to a certain level of organic solidarity in the relevant community? This is a question that might be brought against any statistical theory dealing with large-scale variables—whether it is suicide rates, college admission rates, or rates of alpha decay in a radioactive substance. But the question has a special poignancy when the topic is suicide rates. Here is the reason why.

All the same logical issues of how a statistical (nonuniversal) generalization explains an episode (whether positive or negative for the property in question—decay or suicide) apply. But in the case of an explanation in social sciences such as sociology, the predictions are rarely if ever precise: "More" or "less" is the rule. (Economics tends to generate more precise predictions, though again nowhere near as precise as the theory of spontaneous nuclear radiation.) Being less exact, the social sciences seem to have to work harder to legitimize their statistical theories as explanations.

Durkheim, for instance, has been criticized for committing the so-called ecological fallacy—the error of assuming that a member of a studied population possesses the average characteristics of the entire population as a whole. But the theory of spontaneous radiation is not criticized, although it makes the same move: A sample of plutonium is measured to possess a half-life, and immediately each plutonium atom is assigned the half-life of the sample utilized to make the initial measurement. And this half-life property is invoked in explanation of both individual events of radiation and collective events. The specter of fallacy does not loom, in spite of more recent discussion on the topic of what exactly a statistical "law" is supposed to explain.

What Is Being Explained?

Durkheim was very clear: Sociology is a science of structures. He himself never took individual events—such as incidents of suicide—as the subject of sociology. He was always clear that the science of sociology is a science of institutions and large-scale social facts—and that social structures, such as those

to which sociology drew attention, in turn explain the social facts. Taking Durkheim at his word, we should say that sociology never explains individual events or individual human actions. Sociology explains the differences between different incident rates in different populations. That is all.

But if this is right, then explanation in sociology cannot possibly be construed as falling under the DN model of explanation—at least not in that model's simple form. And if this is right, then it should have some consequences for how we think about statistical laws in physics too: We should not allow the "ecological fallacy" to go unpunished in theories of radioactivity. So in the end, the differences between the sciences—insofar as there are true differences—will serve as important data (as the genus conception of explanation suggests) for more subtle accounts of scientific explanation itself.

Mariam Thalos

See also Causes Versus Reasons in Action Explanation; Common Sense (in the Social Sciences); Covering-Law Model; Durkheim's Philosophy of Social Science; Emergence; Explanation Versus Understanding; Falsifiability; Holism, in the Social Sciences; Hypothetico-Deductivism; Individualism, Methodological; Laws of Nature; *Naturwissenschaften* Versus *Geisteswissenschaften;* Popper's Philosophy of Science; Pseudoscience; Reduction and the Unity of Science; Reductionism in the Social Sciences; Scientific Method; Structural Functionalism, in Social Theory; Weber's *Verstehende* Approach

Further Readings

Collingwood, R. G. (1956). *The idea of history.* Oxford, England: Oxford University Press. (Original work published 1946)

Hempel, C. (1965). Aspects of scientific explanation. In *Aspects of scientific explanation and other essays in the philosophy of science* (pp. 331–496). New York, NY: Free Press.

Kitcher, P. (1981). Explanatory unification. *Philosophy of Science, 48,* 507–531.

Koertge, N. (1992). Explanation and its problems. *British Journal for the Philosophy of Science, 43,* 85–98.

Kyburg, H. (1965). Comment. *Philosophy of Science, 32,* 147–151.

Niiniluoto, I. (1981). Statistical explanation reconsidered. *Synthese, 48*(3), 437–472.

Reichenbach, H. (1951). *The rise of scientific philosophy.* Berkeley: University of California Press.

Reichenbach, H. (1956). *The direction of time.* Berkeley: University of California Press.

Salmon, W. (1971a). Statistical explanation. In W. Salmon (Ed.), *Statistical explanation and statistical relevance* (pp. 29–87). Pittsburgh, PA: University of Pittsburgh Press.

Salmon, W. (Ed.). (1971b). *Statistical explanation and statistical relevance.* Pittsburgh, PA: University of Pittsburgh Press.

Salmon, W. (1984). *Scientific explanation and the causal structure of the world.* Princeton, NJ: Princeton University Press.

Salmon, W. (1989). *Four decades of scientific explanation.* Minneapolis: University of Minnesota Press.

Salmon, W., & Kitcher, P. (Eds.). (1989). *Minnesota studies in the philosophy of science: Vol. 13. Scientific explanation.* Minneapolis: University of Minnesota Press.

Scriven, M. (1962). Explanations, predictions, and laws. In H. Feigl & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science: Vol. 3. Scientific explanation, space, and time* (pp. 170–230). Minneapolis: University of Minnesota Press.

Thalos, M. (2002). Explanation is a genus: On the varieties of scientific explanation. *Synthese, 130,* 317–354.

van Fraassen, B. C. (1980). *The scientific image.* Oxford, England: Clarendon Press.

Weber, M. (1978). *Economy and society.* Berkeley: University of California Press. (Original work published 1922)

Winch, P. (1958). *The idea of a social science.* London, England: Routledge & Kegan Paul.

Explanation Versus Understanding

The distinction between explanation and understanding, taken as two different modes of inquiry, derives from basic features of modern Western philosophy, as articulated chiefly by René Descartes in the 17th century. In his epistemology, Descartes distinguished between an internal and an external realm: mind and body, *cogito* and *res extensa* (extended matter). While the former yielded to philosophical and psychological inspection, material or "natural" phenomena required a "scientific" approach leading to the causal explanation and

prediction of their occurrence. The division was continued in Enlightenment philosophy and, in due course, gave rise to efforts to delineate in ever more precise terms the separation of types of inquiry in the "republic of knowledge."

Naturwissenschaften, Geisteswissenschaften, and *Kulturwissenschaften:* Dilthey and Rickert

During the 19th century, the most prominent distinction was that between the natural and "mental" sciences, the *Naturwissenschaften* and *Geisteswissenschaften,* the former devoted to the explanation of natural processes and the latter to the understanding of "meaning." The most famous proponent of *Geisteswissenschaften* during this period was the philosopher and historian Wilhelm Dilthey (1833–1911). For Dilthey, the recovery of meaning served chiefly as a bulwark to stem the rising tide of "positivist" science. His arguments were directed primarily against British spokesmen of positivist empiricism, such as John Stuart Mill, but one should note some subterranean ties that link him to his opponents. Although an advocate of empiricist methodology for the purposes of sociological inquiry, Mill as a political theorist was also a great defender of liberal individualism; whatever his other preferences, Dilthey shared at least Mill's latter premise, although individualism in his view referred not so much to the ownership of possessions as to the capability of designing cultural artifacts. In elaborating his cultural and historical methodology, Dilthey drew heavily on the teachings of Friedrich Schleiermacher (1768–1834), a distinguished philosopher and theologian of the Romantic period, and also on the writings of the German "historical school." Both Schleiermacher and 19th-century German historians were strongly committed to the notion of *individuality,* construed as a microcosm or as the focal point of a given historical and cultural context.

Dilthey's outlook was molded not only by individualism but also, at least in his early writings, by what has been called *psychologism*. The choice of the term *Geisteswissenschaften* ("mental sciences") was itself an indication that the distinctive feature of the proposed approach was its focus on internal, psychic experience as contrasted to the occurrences of external nature. The basic task of the *Geisteswissenschaften,* which, in Dilthey's view, included history, the humanities, and even part of the social sciences, was to examine manifestations of human creativity and intentionality with the goal of recapturing in past documents and cultural records the original spirit that animated their authors. According to Dilthey, this type of inquiry was bound to produce more reliable and more intelligible results than any other cognitive endeavor, since, in his words, "only what mind has produced, mind can fully understand." The occurrences of the natural or physical world, on the other hand, were relatively opaque and could be rendered accessible only by means of the abstract explanatory constructs of natural science disclosing lawlike uniformities and causal connections: "Nature we explain; psychic life we understand [*verstehen*]."

Obviously, in its reliance on internal experience, Dilthey's approach could serve as a barrier to positivist hegemony only as long as the "mind" and "psyche" were themselves immune from scientific explanation. One of the chief developments during his life, however, was precisely the transformation of psychology into an empirical discipline. On this score, his strategy thus needed to be corrected—and was corrected by, among others, a group of Neo-Kantian thinkers led by Heinrich Rickert (1863–1936), who otherwise were wholly in sympathy with his antipositivist stance. In Rickert's writings, the remedy resided in recourse to the Kantian distinction between facts and (transtemporal) norms; this distinction, he argued, was crucial not only for combating "psychologism" but also for finding an ordering principle in history. Only through their relationship to normative values could historical events acquire cultural significance; for this reason, Dilthey's approach was recast by Rickert under the label of "cultural sciences" (*Kulturwissenschaften*).

Max Weber and Meaningful Action

The bifurcation of the natural and cultural sciences became a cornerstone of Max Weber's (1864–1920) sociology, although one should keep in mind that his methodological views underwent subtle modifications over the years. His early essay on "'Objectivity' in Social Science and Social Policy" (published in 1904) faithfully reflected Rickert's teachings. Sociology was presented by Weber (2011) as one of the cultural sciences, and the latter were defined as "those disciplines which analyze the phenomena of life in terms of their

cultural significance." The significance of cultural events was said to derive from a "value-orientation" toward these events, since "empirical reality becomes 'culture' to us because and insofar as we relate it to value ideas" (p. 76). The examination of social phenomena in terms of cultural meaning, Weber added, was "entirely different from the analysis of reality in terms of laws and general concepts" (p. 77)—although the focus on causal and logical relationships performed an important "preliminary task."

Similar notions can still be found in Weber's monumental study titled *Economy and Society* (published posthumously in 1921), but the gap between cultural understanding and causal analysis was narrowed, and sociology was treated more clearly as a general or systematic science. Weber (1978) defined sociology as "a science which attempts the interpretive understanding of social action" (p. 4), where the term *action* covered "all human behavior when and insofar as the acting individual attaches a subjective meaning to it" (p. 4), while *social* implied that the action "takes account of the behavior of others and is thereby oriented in its course" (p. 4). Meaningful action was segregated in the study from merely externally induced or "reactive behavior" unrelated to an "intended purpose"; but Weber cautioned that the dividing line could not "be sharply drawn empirically" (p. 4). Regarding the notion of action, one should add that it referred only to "the behavior of one or more *individual* human beings" (p. 13)—a carryover of Dilthey's (and Rickert's) individualism. Social aggregates or groupings, in Weber's (1978) view, could never constitute genuine units of analysis.

> For the subjective interpretation of action in sociological work these collectivities must be treated as *solely* the resultants and modes of organization of the particular acts of individual persons, since these alone can be treated as agents in a course of subjectively understandable action. (p. 13)

Weber's Reception

Weber's impact on social science and social theory has been profound and complex. By and large, positivist social scientists have tended to be apprehensive about his notion of interpretive understanding, while many critics of positivism have been attracted to his work precisely because of this feature. Actually, however, reactions on both sides of the fence have been more variegated. Despite his endorsement in some nonpositivist circles, resolute defenders of "cultural" inquiry have found his legacy too deeply imbued with systematic and empiricist leanings to provide an adequate rallying point for their cause. Attitudes in the other camp also have tended to differ. Some champions of a scientific sociology concluded that Weber's work was sufficiently close to their own aspirations to serve as a precedent or springboard for "rigorous" inquiry, given certain corrections or amendments. According to Talcott Parsons (1902–1979), the leading proponent of this view, the major correction needed was the introduction of a general or systematic perspective. While strongly approving the pivotal role assigned to social action, Parsons argued that to permit scientific analysis, human behavior needed to be seen as part of an overarching network or a "social system."

On the whole, Parson's reformulation remained an isolated venture in Weberian scholarship; the bulk of his sympathizers were too apprehensive of "cultural" exegesis to follow his lead. The dominant posture adopted by positivists, and especially by logical empiricists, was to treat Weber indeed as a herald of social research but as one who never caught sight of the promised land. Interpretive understanding in particular—to the extent that it was not completely exorcised—was viewed not as an integral ingredient but as a prefix or embroidery in the anteroom of research, useful for generating "heuristic" insights and for aiding in the formulation of hypotheses. As Otto Neurath (1882–1945), one of the early logical empiricists, once formulated the issue, while helpful for increasing the serendipity or alacrity of the scientist, understanding was as little a part of his actual empirical work as a cup of coffee consumed in the course of his investigations. This "cup-of-coffee" theory of understanding was quickly picked up by other members of the logical-empiricist movement and became a standard weapon in its intellectual arsenal. Typically, the term *understanding* in this context tended to be identified with psychological empathy or with the reenactment or "reproduction" of mental and emotive processes.

Analytic Philosophy Versus Continental Philosophy

Battles sometimes continue to be fought when the original motives have vanished or been seriously

modified. While the empiricist offensive against Weber's legacy was still in full swing, developments were already afoot that were destined to recast the underlying issues in completely new terms. These developments occurred in two very different settings and involved very divergent lines of argument: One arose within the largely Anglo-American confines of the "analytical philosophy of language"; the other emerged on the Continent, under labels such as *phenomenology* and *existential analysis*. Despite their heterogeneity, the two developments moved toward (at least partial) convergence with regard to the role of interpretive understanding; using a shorthand formula, one might say that both initiatives threw a new light not only on the much belabored issue of "psychologism" but also on the much more weighty and age-old heritage of methodological "individualism" (still prominent in Weber's case).

Analytic Philosophy and the Linguistic Turn: Wittgenstein and Winch

In the analytical context, the turn toward language signified first of all a realization that empirical reality could not be directly grasped without an adequate conceptual and linguistic framework; in the long run, however, the same turn carried with it another implication important for social inquiry: the insight that human action is social from the beginning, since the meaning of an action cannot be articulated even by the individual actor without recourse to language, that is, a shared pool of significations. Initially, it is true, language analysts—like their positivist companions—shunned all reference to purposive meaning, whether individual or social, and to the enterprise of interpretive understanding. The contours of a rapprochement emerged only in the later writings of Ludwig Wittgenstein (1889–1951), especially in his emphasis on ordinary language and the notion of "language-games" embedded in commonsense conventions. Once linguistic practices were seen as intimately "interwoven" with concrete "life-forms" and worldviews, the feasibility of a "cultural" interpretation became apparent. Nevertheless, Wittgenstein's own attitude in this matter remained ambivalent to the end, as he left open (or failed to block) the road to an empiricist treatment of language and the reduction of meaning to behavior.

Whatever Wittgenstein's preferences may have been, a number of followers—foremost among them being Peter Winch (1926–1997)—have developed the notion of language-games into a springboard for cultural and social analysis. Winch's classic *The Idea of a Social Science* (2008) placed decisive stress on the task of interpretive exegesis, or on "the central role which the concept of understanding plays in the activities which are characteristic of human societies" (p. 21). As the study tried to show, however, the interpretation of social behavior could not rely solely on the intentions of an individual actor, since these intentions were intelligible only within a language community and in terms of the rules of conduct and language usage operative in that community. Social analysis thus was predicated on the understanding of intersubjective standards: "It is only a situation in which it makes sense to suppose that somebody else could in principle discover the rule which I am following that I can be said to follow a rule at all" (p. 28). Basically, Winch argued that social understanding involves "grasping the *point* or *meaning* of what is being done or said" and thus an effort "far removed from the world of statistics and causal laws" and "closer to the realm of discourse" (p. 108).

The aspect of Winch's (2008) study that occasioned the most lively controversy among reviewers was the thesis of the uniqueness and virtual incommensurability of individual language-games and life-forms. Viewing life-forms as closely knit webs of opinions and ideas, Winch considered it "nonsensical to take several systems of ideas, find an element in each which can be expressed in the same verbal form, and then claim to have discovered an idea which is common to all the systems"; rather, the "very nature" of human society was "to consist in different and competing ways of life, each offering a different account of the intelligibility of things" (p. 103). His argument on this score has been criticized from many quarters—most vocally by adepts of a "critical positivism" committed to the proposition that scientific research, although fallible and in need of constant revision, is destined ultimately to grasp objective reality independently of historical conditions and cultural variations.

The Continental Approach: Husserl, Heidegger, and Schütz

The argument was also challenged from the side of Continental thought. In that context, phenomenology was to some extent a continuation of Dilthey's

and Rickert's efforts to elucidate the dimension of "meaning" as a counterpart to scientific explanation. Paralleling Rickert's initiative, Edmund Husserl's (1859–1938) early writings launched a broad-scale attack on "psychologism" and on the reduction of thought to empirical processes; by comparison with Neo-Kantianism, however, Husserl from the beginning extended the scope of inquiry beyond normative values to the full range of (natural and cultural) phenomena amenable to human cognition. While reformulating and sharpening the insights of his predecessors, Husserl remained at least in one respect heir to their perspective: in the attachment to individualism or to an individual-egological "consciousness." At least in this respect, his approach replicated the solipsistic dilemma of early language analysis and of much of modern philosophy: To the extent that consciousness was presented as the "transcendental limit" of the world, the domain of intersubjective understanding and clarification of meaning was obliterated. In his later writings, Husserl sought to overcome this dilemma by introducing the notion of the "life-world," or the world of mundane experience, but the relationship between mundanity and consciousness was never fully clarified.

By that time, the thrust of phenomenological analysis had already been profoundly reshaped by Martin Heidegger's (1889–1976) delineation, in *Being and Time,* of a "hermeneutical phenomenology" or "existential ontology." Understanding of meaning, in Heidegger's conception, was no longer the prerogative of individual cognition or consciousness but a basic attribute of man's existential condition or *Dasein,* construed as "being-in-the-world." Far from occupying the role of an alien spectator, *Dasein* was seen as enmeshed in a fabric of "pre-understanding" and "pre-predicative" experience, a fabric that was basically intersubjective and cultural in character. As it happened, neither Husserl nor Heidegger was concerned with articulating the implications of their thoughts for social inquiry. The elaboration of a "phenomenological sociology" was chiefly the accomplishment of one of Husserl's students, Alfred Schütz (1899–1959). In one of his first major works, *The Phenomenology of the Social World,* Schütz tried to effect a merger of Weberian sociology and Husserl's teachings by tracing the notion of meaningful social action to an underlying stratum or stream of constitutive consciousness. In his later writings, he grew steadily weary of egological methodology and was increasingly preoccupied with the domain of the "life-world," the domain of "common sense" and everyday activity. Drawing, at least in part, on Heideggerian insights, he treated "understanding" not simply as an individual cognitive faculty but as a multidimensional category denoting an existential or "experiential form of common-sense knowledge of human affairs" as well as an epistemological problem and a "method peculiar to the social sciences."

Recent Developments: Habermas, Apel, Gadamer, and Ricoeur

The tension between invariance and contingency, individualism and intersubjectivity are not the only predicaments besetting social inquiry today. Conjoined with these issues, the sketched skirmishes and developments have engendered a quandary that touches the core of cognition and the basic structure of the republic of letters: the quandary regarding the relationship between science and understanding, knowledge and self-knowledge. Although universally noted, the dilemma has been debated in recent years with particular intensity by German and French thinkers; preponderantly, these debates have centered on the proper range or scope of interpretive exegesis. One argumentative strategy, favored by representatives of the "Neo-Frankfurt School" of social research, has been to maintain the integrity of traditional types of inquiry while resisting the tendency toward segregation or one-sided hegemony. Thus, dedicated to the long-range prospect of human self-understanding and emancipation but disturbed by the quasi-natural constraints of modern society, Jürgen Habermas (1929–) and Karl-Otto Apel (1922–) have advocated the juxtaposition of hermeneutics and scientific analysis (and also their combination for purposes of critical social inquiry). On the other hand, hermeneuticists influenced by Heidegger have tended to redefine interpretive exegesis by deemphasizing the aspect of subjective purpose and intentionality; in his *Truth and Method,* Hans-Georg Gadamer (1900–2002) presented history not so much as an emanation of individual practice but as a complex learning process in which "man" is able to decipher himself only through encounters with a great variety of cultural practices and

institutions. Seen in this light, the scope of hermeneutical understanding is bound to be broad and devoid of clear-cut boundary lines. In the French context, Paul Ricoeur (1913–2006) has argued in favor not only of the coexistence but also of an intimate reciprocal dialogue and symbiosis of understanding and scientific explanation.

Drawing mainly on Gadamer's and Ricoeur's hermeneutics, some political theorists have recently launched the project of a "comparative political theory" where understanding (coupled with some explanation of social background) aims to transgress national and cultural boundaries in the direction of a cosmopolitan mode of discourse.

Fred Dallmayr

See also Agency; Being-in-the-World; Causation in the Social Sciences; Causes Versus Reasons in Action Explanation; Ethnomethodology; Existential Phenomenology and the Social Sciences; Explanation, Theories of; Frankfurt School and Critical Social Theory; Hermeneutics, Phenomenology, and Meaning; Holism in Social Sciences; Language-Games and Forms of Life; Life-World; Logical Positivism/Logical Empiricism; *Naturwissenschaften* Versus *Geisteswissenschaften*; Relativism and the Social Sciences: From the Sapir-Whorf Hypothesis to Peter Winch; Rule Following; Transcendental Pragmatics; Weber and Social Science: Methodological Precepts; Weber's *Verstehende* Approach

Further Readings

Brown, R. R. (1963). *Explanation in social science.* Chicago, IL: Aldine.

Dallmayr, F., & McCarthy, T. A. (Eds.). (1977). *Understanding and social inquiry.* Notre Dame, IN: University of Notre Dame.

Hodges, H. A. (1952). *The philosophy of Wilhelm Dilthey.* New York, NY: Humanities Press.

Manninen, J., & Tuomela, R. (Eds.). (1975). *Essays on explanation and understanding.* Dordrecht, Germany: Reidel.

Radintzky, G. (1968). *Contemporary schools of metascience* (2 vols.). Göteborg, Sweden: Akademieförlaget.

Taylor, C. (1964). *The explanation of behavior.* London, England: Routledge & Kegan Paul.

Toulmin, S. (1961). *Foresight and understanding.* London, England: Hutchinson.

Von Wright, G. H. (1971). *Explanation and understanding.* Ithaca, NY: Cornell University Press.

Weber, M. (1978). *Economy and society* (G. Roth & C. Wittich, Eds.). Berkeley: University of California Press. (Original work published 1922 in German as *Wirtschaft und Gesellschaft,* Tubingen, Germany: Mohr)

Weber, M. (2011). *Methodology of social sciences* (E. A. Shils & H. A. Finch, Eds. & Trans.). New Brunswick, NJ: Transaction Books. (Original work published 1949

Winch, P. (2008). *The idea of a social science and its relation to philosophy.* New York, NY: Routledge. (First edition, 1958; second edition, 1990)

Ziff, P. (1972). *Understanding understanding.* Ithaca, NY: Cornell University Press.

FALSIFIABILITY

This entry gives an account of *falsifiability* both as championed in particular by Karl Popper and also more generally and examines its wider implications for scientific methodology. It may be useful to start by pointing to the connection between falsifying in science, as a desideratum for choosing scientific theories, and the related logical argument form of *modus tollens: p→q/¬q/therefore ¬p*.

Falsifiability—or more accurately, *empirical falsifiability*—is a central notion in Karl Popper's philosophy of science, and more particularly in his strategy for *demarcating* science from *pseudoscience*. Roughly, a theory is (empirically) falsifiable if it is refutable by experience. More accurately, we might say that a theory is empirically falsifiable if there is a possible observation statement with which it is logically inconsistent. Consider the theory "All rabbits are brown" and the statement "There is a black rabbit." These cannot simultaneously be true; hence, the former must be classified as false if the latter is classified as true. Now we need only to add that "There is a black rabbit" is a possible *observation* statement—a statement that might be made on the basis of sensory experience—in order to conclude that "All rabbits are brown" is *empirically* falsifiable.

Here, however, we reach the first hurdle. It is dubious that existential statements such as "There is a black rabbit" are direct, or pure, statements of what we observe. Rather, it would seem that two people might have the same sensory inputs at a point in time but disagree on what those inputs show to be true *in virtue of having different theories*. I may be convinced that a black-looking rabbit has just had its fur dyed, after discovering a fiendish plot by a rival scientist to lead my investigations into lagomorphs astray. But another person, upon seeing the same thing at the same time, might take the rabbit to be genuinely black. And fictional tales aside, the history of science shows the actual methodological significance of the *theory-laden* nature of observation. For example, Galileo observed Neptune more than 200 years before its recognized "discovery" (which we will discuss later), but he mistook it for a star.

On the face of it, this recognition is not too problematic. What's observable becomes a matter of theoretical context, as well as physical context, and what's falsifiable becomes similarly contextual as a result. However, to demand that all scientific theories be falsifiable then becomes troublesome—at least, that is, if we require that our observation statements should only ever be laden with *scientific* theories. Imagine that observation statement *O* laden with theory *T* conflicts with theory *T*. If we are to take *O* seriously, must we demand that *T* be falsifiable, in turn? If so, then the observations that could falsify it will also be theory-laden. Thus, we should demand that those further theories be falsifiable too. And so on, back in an infinite regress or around in a circle. Popper's response to this kind of difficulty was to say that there is no firm place to stop and that we can always go on testing. One problem with this view, for those inclined to a realist view of science (where science makes continual progress toward the truth), is that it is hard to see how stopping in one

place rather than another is anything more than a pragmatic matter.

But what if—against the views of Popper and Thomas Kuhn—some observations are not theory-laden? For the sake of argument, let's imagine for the remainder of this discussion that *no* observations are theory-laden. Is it then correct to say that all scientific theories should be empirically falsifiable?

Unfortunately it would seem not, because most scientific theories, if not all scientific theories, predict nothing whatsoever *when taken in isolation*. As Pierre Duhem noted, writing about physics, experiments cannot condemn isolated hypotheses. Consider, for example, Newton's first law of motion: A body remains at constant velocity (or rest) unless a resultant force is acting on it. Nothing in this statement tells us what kinds of forces there are or how to determine when they are present. So the law does not predict what any actual body will do in any actual situation. The upshot is that an observation of an accelerating body could only ever falsify Newton's first law *in combination* with an additional *theoretical* account of what forces there are, how to measure their magnitudes in particular instances, and so forth.

A natural response is to say that this law is part of an empirically falsifiable *system* of theories; Popper suggested as much, albeit rather briefly. So might one then suggest that science contains such systems of theories, whereas pseudoscience does not? Again, there are serious difficulties. If the claim is to be that *any* discipline involving empirically falsifiable systems of theories is scientific, then this is consistent with highly dogmatic, and inflexible, rules for theory selection/preference. To see this, it will help to consider the methodological consequences of Duhem's aforementioned thesis. What should we do when faced with an observation statement that falsifies some system of theories? For all that logic tells us, it is always acceptable to preserve a favored component of the system and apportion blame elsewhere. This is illustrated beautifully by the story of Urbain Leverrier's investigations of planetary orbits.

Using Newtonian mechanics and hypotheses concerning the orbits of the known planets—note that these orbits had a theoretical character, insofar as only the positions of the known planets at particular times were directly measured—it was possible to predict the orbit of Uranus. Yet observations indicated that Uranus did not have the predicted orbit. This did not, however, shake Leverrier's belief in Newtonian mechanics. Instead, he thought that the problem lay with the other, auxiliary, information used to derive the prediction. In particular, he considered the possibility that the information was incomplete and that there was a previously unknown planet exerting gravitational influence on Uranus. Hence—and do not underestimate the mathematical difficulty of this task or the considerable guesswork involved (because, e.g., a hypothetical mass for the planet had to be selected)—he was able to suggest a place to look. And sure enough, Neptune was found.

Several years later, it was noted that the orbit of Mercury was similarly aberrant; that is to say, Mercury did not move in the path predicted by Newtonian mechanics and the available auxiliary hypotheses concerning other bodies in the solar system. Unsurprisingly, Leverrier sought to repeat his previous trick. He predicted a planet Vulcan. But Vulcan does not exist. And we now know that the orbit of Mercury can be explained by relativity theory.

The point behind the story is that the circumstances in these two scenarios were highly similar, but in one case a central physical theory was retained and in the other it was abandoned. But Newtonian mechanics need not have been abandoned *on logical grounds*. When Vulcan was not found, indeed various different possibilities were examined—the existence of multiple unknown planets, of other massive bodies such as asteroid belts, and so forth. And no matter how many such possibilities were ruled out, countless many—infinitely many, some philosophers would urge—remained. Multiple invisible bodies might even have been posited. And before you dismiss this as crazy, consider the "dark matter" posit of more recent astrophysics.

The lesson to draw from the story is that to adopt any old falsifiable system is not to do science. Rather, what one does and what *one is willing to do* with the given system is crucial. (Alternatively, one may think at the level of science itself and of the functions *it* fulfills.) In many pseudosciences, some hypotheses are never questioned or held open to revision. Think, for example, of the way astrologists merely accommodated the existence of Neptune, rather than questioning the notion that planetary positions have some effect on our lives.

In closing, one further difficulty with requiring falsifiability of scientific theories, which is independent of the prior considerations, is worthy of note. This concerns *probabilistic* laws or theories, such as those in quantum mechanics. Consider the hypothesis that a given coin is fair—that is, that the

probability of a heads result is equal to the probability of a tails result when it is flipped. Strictly speaking, this is compatible with an observation statement of the form "The coin has landed on heads on the last *n* flips," where *n* is any natural number. Yet it would not be unreasonable to abandon the theory that the coin was fair if *n* were sufficiently high.

Darrell P. Rowbottom

See also Critical Rationalism; Duhem-Quine Thesis and the Social Sciences; Given, Myth of the; Observation and Theory-Ladenness; Popper's Philosophy of Science; Pseudoscience

Further Readings

Duhem, P. M. M. (1954). *The aim and structure of physical theory* (P. P. Weiner, Trans.). Princeton, NJ: Princeton University Press. (First English edition published 1906)

Gillies, D. (1971). A falsifying rule for probability statements. *British Journal for the Philosophy of Science, 22,* 231–261.

Popper, K. R. (1959). *The logic of scientific discovery.* New York, NY: Basic Books.

Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review, 60,* 20–43.

Rowbottom, D. P. (2010). Corroboration and auxiliary hypotheses: Duhem's thesis revisited. *Synthese, 177,* 139–149.

Feedback Mechanisms and Self-Regulatory Processes in the Social Sciences

This entry introduces the notions of *feedback* and *self-regulation,* together with their variations used in mechanisms employing loops as analytical tools of modeling social phenomena in systems theory. It explains the importance of such interactive processes and why such systems thinking is ubiquitous in social science; the entry also offers a number of helpful examples.

Background

Beginning in the 1940s, the engineer's concept of *feedback* entered the social sciences. The essence of the concept is a circle of interactions—a *closed loop* of action and information. Something, presumably causal pressures or information, "feeds around" a hypothesized loop of interactions and eventually "feeds back" to its point of origin. Such loops are essential for automatic control devices ranging from Watts's centrifugal governor for a steam engine to modern thermostats, toilet tank fill valves, and automobile cruise controls. But the thinking that underlies the notion of feedback loops is not simply an engineering idea but rather a much more general and pervasive perspective with a rich history in the social sciences.

The notion traces back implicitly at least 250 years in social science literature, and more than 2,000 years more generally. When the loop character of societal reasoning has been explicit, it has been variously characterized as mutual or circular causality, homeostasis, control, recursion, self-reference, reciprocal relations, interdependence, and the like. Thus, feedback loops, by whatever name they are known, are at the heart of the social and policy sciences. It is not an exaggeration to say that great social scientists are feedback thinkers and great social theories are feedback thoughts.

Feedback and Self-Regulation

Self-regulation is the tendency of an individual or group to try to hold itself in some preferred condition. The generic idea postulates a *goal* of some sort and a perception that may align with the goal or create a discrepancy. A discrepancy between the perceived and preferred states creates pressures to take some actions to close the gap and bring the system back into some balance. Figure 1 captures the general idea.

Early terminology called such loops *negative* feedback loops because tracing around a change in such a loop produces pressures that try to counter or negate the change. Current terminology calls such loops *balancing* or *counteracting*.

Examples abound for such *balancing* or *counteracting* loops. If one is too cold (a gap between desired and actual conditions), one puts on a sweater or coat, thus working to close the gap and be comfortable. If the supply of some commodity is insufficient to satisfy the demand, prices would tend to rise, pulling the demand back closer to the supply (one of Adam Smith's "invisible hands" at work). If a population outstrips its ability to feed itself, various mechanisms would act to limit growth (if the food supply can't be increased fast enough, Thomas Malthus saw socially conscious "preventative" checks and nasty, unavoidable "positive"

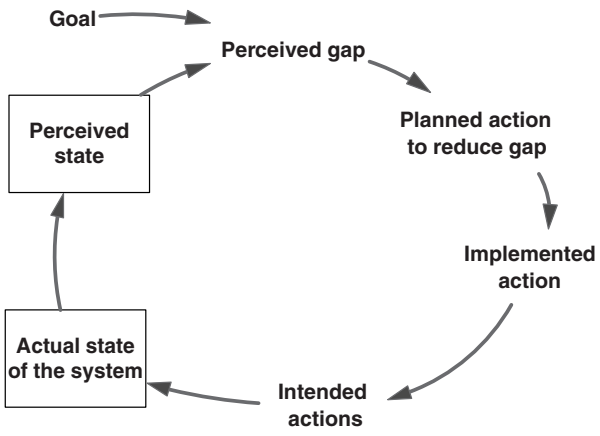


Figure 1 Generic Self-Regulating Feedback Loop: A Goal-Seeking "Balancing" Loop for an Individual or Group

Source: Author.

checks, both of which create balancing loops). If government excesses emerge in a democracy, the self-interests of the governed should produce actions that would control the excesses (the authors of the *Federalist Papers* explicitly argued for such self-regulatory governmental mechanisms in the U.S. Constitution). More modern social and behavioral science examples include Leon Festinger's notion of the motivating power of "cognitive dissonance"; the importance of vision, mission, and monitoring to guide planned action in strategic management; and the actions of the Open Market Committee of the U.S. Federal Reserve to adjust interest rates to try to keep inflation in check and the economy on a stable growth path.

Self-Regulatory Loops in Context

Suffice it to say, all *purposeful* self-regulatory action of an individual or group can be thought of in terms of one or more balancing feedback loops of the sort generalized in Figure 1 (sometimes without an explicit identified goal). In practice, things are much more complex than that single loop, for two reasons.

First, in any social, economic, or political setting we have many actors, not just one, and thus many goals and many goal-seeking balancing loops interacting over the same issues. The picture becomes complex very quickly. Second, in any one of those balancing feedback loops, there are often implicit goals that people have not expressed, even to themselves; actions often have unintended consequences; and while efforts focus on closing some perceived gap between goal and perception, the system can move of its own accord under a host of other forces, including explicit resistance to implemented goal-directed actions. Figure 2 suggests this more complicated picture.

The diagram in Figure 2 contains six feedback loops. With more actors or stakeholders and more interconnected sets of loops as in Figure 2, the interacting complexity grows dramatically, and we have the complicated dynamics that social, economic, and policy sciences struggle to understand. Furthermore, some of the loops in Figure 2, or a more complicated version with more actors, may not have the balancing or self-controlling character of the loop in Figure 1. Some of those loops could be so-called positive or reinforcing feedback loops.

Reinforcing Feedback Loops

Unlike balancing loops (the aforementioned *negative feedback loops*), which tend to counter a change, *reinforcing loops* tend to move further in the direction of change and amplify it, adding instability.

Again, examples abound, many of which were recognized in the social science literature before the emergence of the engineers' uses of feedback: John Stuart Mill noted that speculators, seeing a rise in the price of a stock or commodity, come to expect the rise to continue and so increase their purchases, thus driving up the price still further. In a *vicious circle,* a term that in the 1600s flagged faulty logic, we now have a bad situation that leads to its own worsening. And when supporters of some social or political movement grow and attract still more supporters simply because of growing popularity, we have what's called the "bandwagon" effect, named for the tendency of people to be attracted to crowds following the first vehicle in a circus parade. The *self-fulfilling prophecy* was so labeled by Robert K. Merton to flag a prediction that starts out false but becomes true because people come to believe it. Merton's examples were predictions of the insolvency of a bank, the fear of failing a test, and the prejudices that discriminate and limit the accomplishments of a group or race and thus reinforce the prejudgments.

Balancing and Reinforcing Loops Together

The importance of self-regulating and self-reinforcing feedback loops in social science stems from what we can understand by employing a feedback perspective. Whether stated, implicit, or unrealized, interconnected balancing and reinforcing

feedback loops are at the heart of our deepest social and economic theories and understandings. But this takes a complex view. Figure 3, for example, shows a glimpse of the complex interactions of feedback loops that computer simulation shows us can by themselves generate the patterns of growth, stagnation, and decay we observe in our urban centers.

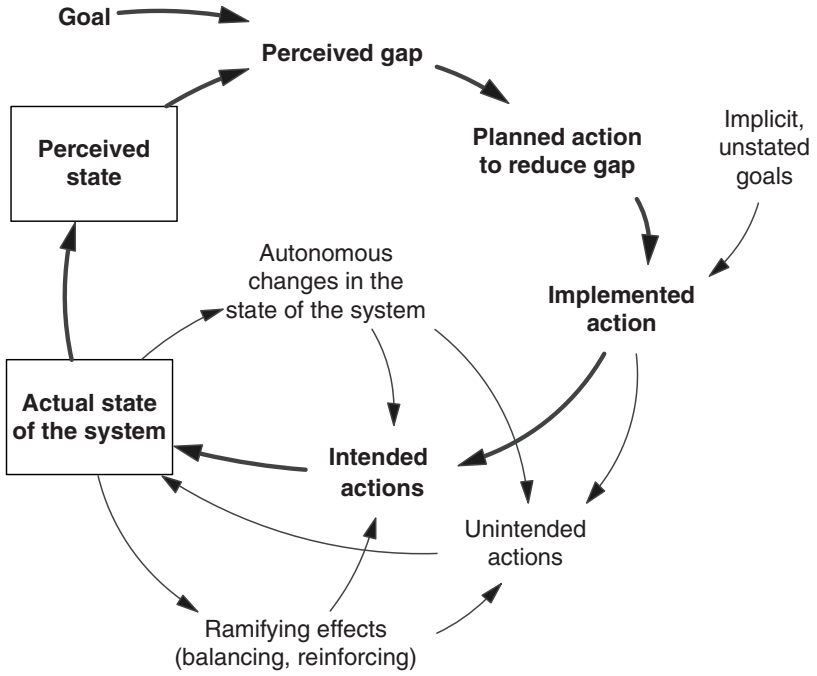


Figure 2 Goal-Seeking Balancing Loop of Figure 1 Complicated by Implicit Goals, Unintended Consequences, and Systemic Changes That Can, and Often Do, Thwart Attempts at Self-Regulation

Source: Author.

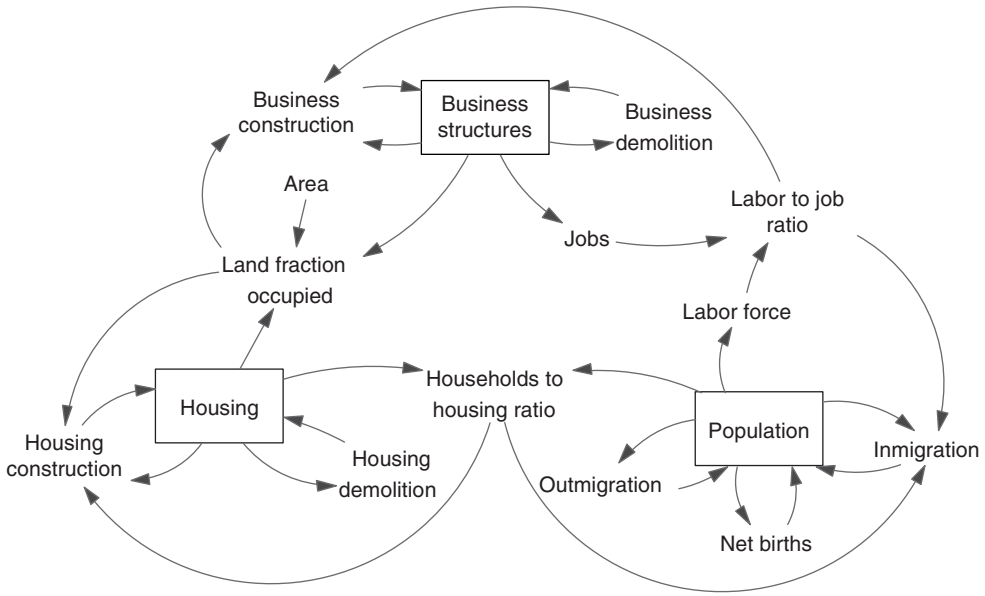


Figure 3 Self-Regulating and Self-Reinforcing Feedback Loops in the Dynamics of a City, Creating an Endogenous Theory of Urban Growth, Stagnation, and Decline

Source: Author.

The importance of a feedback perspective in the philosophy of social science rests on the endogenous point of view exemplified in Figure 3. Our strongest social science theories are *endogenous theories*. Feedback loops enable the endogenous point of view and give it structure.

George P. Richardson

See also Complex Networks Theory and Social Phenomena; Complexity; Goal-Directedness; Invisible Hand Explanations; Mathematical Models, Use in the Social Sciences; Prophecy, Self-Fulfilling/Self-Defeating; Retrodiction and the Epistemology of Future Studies; Systems Theory

Further Readings

Jervis, R. (1997). *System effects: Complexity in political and social life*. Princeton, NJ: Princeton University Press.

Maani, K. E., & Cavana, R. Y. (2007). *Systems thinking, system dynamics: Understanding change and complexity*. Auckland, New Zealand: Prentice Hall.

Richardson, G. P. (1991). *Feedback thought in social science and systems theory*. Philadelphia: University of Pennsylvania Press.

Feminism: Schools of Thought

Feminist schools of thought are many and diverse. However, in one way or another, all of them are concerned with improving women's status, autonomy, and/or socioeconomic circumstances. Furthermore, almost all schools of feminist thought examine gender relations in the public and private realms and/or attempt to rehabilitate characteristics traditionally associated with women. Among the main schools of feminist thought are the following: liberal, radical (libertarian and/or cultural), Marxist-socialist, psychoanalytic and care focused, existentialist and postmodern, multicultural and global, ecofeminist, and third wave. Although some feminist schools are highly controversial and/or daunting to the average person, they are all worthy of consideration. Taken together, they provide new perspectives on knowledge, reality, and social justice.

Liberal Feminism

Liberal feminism has its roots in English philosophers such as Mary Wollstonecraft (1759–1797), John Stuart Mill (1806–1873), and Harriet Taylor Mill (1807–1858). It also traces its beginnings to the women's suffrage movement, especially in the United States. In the 19th century, liberal feminists like Mary Wollstonecraft emphasized that if women were to receive men's education, women would stop behaving like children or, worse, caged birds, potted plants, or ornamental objects for men. Agreeing with Wollstonecraft, John Stuart Mill and Harriet Taylor Mill emphasized that women's autonomy also necessitated women's movement into the public sphere (particularly the paid workforce). The Millses also pushed women's suffrage as a corrective for women's subordination to men. Armed with the vote, women could express their interests and values, thereby gaining power in society. Suffrage did not come easy for women, however. In the United States, for example, it took about a half-century for the passage of the 19th Amendment to the U.S. Constitution, which enfranchised women.

Like the Millses, 20th-century liberal feminists claimed that women's subordination to men is predicated on a set of assumptions about women's gender identity and roles. These assumptions are used first to justify women's near confinement to the private realm, where they are expected to perform most of the domestic chores and care of dependents, and second to limit women's access to the public realm (advanced education, the political arena, the marketplace, and the high-status and high-paid professions). Largely, through the efforts of 20th-century feminists and contemporary organizations like the National Organization of Women (NOW), women in the United States, for example, are now comfortable in the public realm, working as chief executive officers, legislators, professors, physicians, lawyers, and so forth. Still, NOW's 1970s Equal Rights Amendment has not been ratified by enough states, and women still do most of the household's domestic and care work. For example, a 2002 survey found that women do 70.6 hours of domestic work a week, whereas men do 37.3 hours of domestic work a week. Similarly, women do 9.1 hours of elder care work per week, whereas men do 5.7.

Radical Feminism

Radical feminists alleged that the liberal feminist agenda is a weak program of reform rather than a strong revolutionary movement aimed at the heart of patriarchy. They claimed that patriarchy—a state

of affairs in which male "fathers" rule—is characterized by competition, hierarchy, and brute power. According to radical feminists, patriarchy cannot be reformed; it needs instead to be destroyed. It is not just patriarchy's legal, political, and socioeconomic structures that must be overhauled, its sexual and cultural institutions (especially the family) must also be reconceived.

There are two types of radical feminists: (1) radical-libertarian feminists and (2) radical-cultural feminists. Radical-libertarian feminists rejected the notion that the only kind of sex that is "good" for women is to be found in a long-term love relationship between a married heterosexual couple. Instead, they claimed that any kind of sex that gives women pleasure is good: sex between men and women, sex between women, sex with both men and women, autoeroticism, sadomasochistic sex, and even intergenerational sex. In addition, radical-libertarian feminists put forward another controversial view—namely, that natural reproduction be replaced by technological reproduction. They thought that thinkers such as Shulamith Firestone were correct to argue that no matter how much educational, political, and economic equality women achieve, they will not be liberated from men unless they stop bearing children. Radical-libertarian feminists also applauded when Firestone claimed that it would no longer be necessary for men to display only masculine identities and behaviors and for women to display only feminine ones. Instead, men and women would become encouraged to mix and match feminine and masculine characteristics in any way they wished. As a result of this new choice, not only men and women but also all of culture would be androgynous. Furthermore, as this new androgynous culture developed, the categories of the technological and the aesthetic, together with the categories of the masculine and feminine, would self-destruct.

Disagreeing with radical-libertarian feminists about both sex and reproduction, radical-cultural feminists claimed that sex in general, but certainly sex between women and men, was inherently dangerous. Indeed, they said that as it has been experienced so far, heterosexuality is a "compulsory institution" that justifies male sexual abuse of women representatively in pornography and actually through the use of prostitutes, sexual harassment, rape, and domestic violence. Radical-cultural feminists therefore encouraged women to tap into their own sexuality, a sexuality not created in service of men's desires, and to explore the "lesbian" within themselves.

Equally as suspicious of artificial reproduction as of institutionalized heterosexuality, radical-cultural feminists urged women to view high-tech infertility services and plans for an artificial womb not as procreative options for women but as a means for men to seize control of women's reproductive power. Radical-cultural feminists stressed women's need to safeguard their ability to bring life into the world through their own bodies, for without this life-giving power, men would have more leverage over women than ever before.

Marxist Feminism

Marxist-socialist feminists criticized both liberal and radical feminists for neglecting women's class status. Although it is true that women in capitalist patriarchies occupy a subordinate position relative to men on account of their sex, this does not mean that all women are equally oppressed. Specifically, it is worse being a poor woman than a rich woman in a capitalist patriarchy.

Marxist-socialist feminists realized that it is somewhat contradictory to be both a Marxist-socialist and a feminist at one and the same time. Should class be used as the primary lens through which to interpret women's (and some men's) oppression? Or should sex be the primary lens, with class playing a secondary role to it? Or was there some way to make class and sex equally important in interpreting women's oppression?

According to the feminist thinker Chris Beasley, Marxist-socialist feminists developed three different approaches to the questions just raised, regarding three ways of seeing the relationship between the two systems of oppression (class and sex)—that is, the relationship between an economic or political analysis in terms of "class" and a feminist-psychological one of "sex." Beasley labeled these three approaches (1) the two-tier or dual-systems approach, (2) the unified-system or capitalist-patriarchy approach, and (3) the "dynamic-duo" approach. Ultimately, none of these approaches to mesh Marxist-socialism with feminism has been entirely successful. Class and sex remain competitors for the soul of Marxist-socialist feminism, with class being given a greater or lesser edge over sex. Although Marxist-socialist feminists easily agree that women's low socioeconomic status is multiply determined by women's role

not only in production but also in reproduction, the socialization of children, and sexual relationships, most of them feel that women's place, or lack of place, in the paid workforce best explains women's low socioeconomic status.

Psychoanalytic and Care-Focused Feminists

To the degree that liberal, radical, and Marxist-socialist feminists focus on the macrocosm (patriarchy or capitalism) in their respective explanations of women's oppression, psychoanalytic and care-focused feminists look to the microcosm of the individual, claiming that the roots of women's oppression are embedded deep in the psyche. Some psychoanalytic feminists relied on Freudian constructs such as the pre-Oedipal stage, during which all infants are symbiotically attached to their mothers, whom they perceive as omnipotent, and the Oedipal stage, during which boys much more successfully than girls are integrated into society by giving up their first love object, the mother. In contrast, other psychoanalytic feminists relied on Lacanian constructs such as the *symbolic order* (society's language system) and the *prelinguistic imaginary.* In either event, psychoanalytic feminists claimed that gender identity is rooted in the experiences infants/children have during these stages. These experiences, most of which we scarcely remember, are supposedly the cause of individuals viewing themselves as males or females, as masculine or feminine. Moreover, these same experiences are the cause of society's elevating masculine values, accomplishments, and desires over feminine values, accomplishments, and desires. Hypothesizing that in a nonpatriarchal society where, for example, men and women "dual parented" their children, masculinity and femininity would be both differently constructed and valued, psychoanalytic feminists recommended that we work toward such a society by changing our early infantile experiences or, even something more radical, by transforming the linguistic structures that cause us to think of ourselves as men or women.

Like psychoanalytic feminists, care-focused feminists probed women's (and men's) psyches, trying to find there more information about society's privileging of things masculine over things feminine. Care-focused feminists stress that at least in Western societies women as a group are associated with values, virtues, and characteristics such as interdependence, community, connection, emotion, body, nature, and care. In contrast, men as a group are associated with values, virtues, and characteristics such as independence, autonomy, separateness, reason, mind, culture, and justice. Care-focused feminists explain these associations in a variety of ways, attending to the ways in which societies systematically shape men's and women's distinct identities and behaviors. But whatever explanation they offered, care-focused feminists regarded women's capacities for care as a human strength rather than a human weakness. Moreover, care-focused feminists spent considerable time developing a feminist ethics of care as a complement to, or even a substitute for, a traditional ethics of justice. As they saw it, were men to do as much caring work as women do, women's status would rise.

Existential and Postmodern Feminists

Looking into women's psyches more deeply than even psychoanalytic and care-focused feminists, Simone de Beauvoir provided an ontological-existential explanation for women's oppression. In *The Second Sex,* an authoritative text of 20th-century feminism, she argued that woman is oppressed by virtue of her otherness. Woman is the other because she is not man. Man is the free, self-determining being who defines the meaning of his existence; woman is the other, the object whose meaning is determined for her. If woman is to become a self, a subject, she must, like man, transcend the definitions, labels, and essences limiting her existence. She must make herself be whatever she wants to be.

Postmodern feminists took de Beauvoir's understanding of otherness and turned it on its head. Woman is still the other, but rather than interpreting this condition as something to be rejected, postmodern feminists embraced it. They claimed that woman's otherness enables individual women to stand back and criticize the norms, values, and practices that the dominant male culture (patriarchy) seeks to impose on everyone, particularly those who live on its periphery. Thus, otherness, for all its associations with being excluded, shunned, rejected, unwanted, abandoned, and marginalized, has its advantages. It is a way of existing that allows for change and difference. Women are not unitary selves, essences to be defined and then ossified. On the contrary, women are free spirits.

Multicultural and Global Feminism

Multicultural and global feminists challenged "female essentialism"—that is, the view that "Woman" exists as some sort of Platonic Form. They also confronted "female chauvinism"—that is, the tendency of some women, privileged on account of their race or class or some other characteristics about themselves, to presume to speak on behalf of all women. Thus, it is not surprising that both multicultural and global feminists insist on highlighting women's differences from each other.

In the United States, as in other multiracial and multiethnic societies, multicultural feminists stressed the inseparability of the structures and systems of gender, race, and class. They claimed it was wrongheaded for women to focus exclusively on their oppression as women, as workers, or as members of a certain racial or ethnic group. Instead, they claimed that each woman, or each relatively distinct group of women, needs to understand how everything about her—the color of her skin, the amount of money in her purse, the condition of her body, the object(s) of her sexual desire, the date on her birth certificate—provide part of the explanation for her status as a woman. Women are jeopardized in multiple ways, and systems of oppression are interlocking. Racism, sexism, and classism are never separable in fact, even if they are separable in theory, and there is no way to eliminate one of these isms while not eliminating the others. Thus, the fundamental task of multicultural feminists is to fight against all oppressive forces, including one's own self-doubting and self-denigration.

Agreeing with multicultural feminists that the definition of feminism needs to include class and race as well as sex, global feminists emphasized that being an oppressed woman in a developing nation or an authoritarian regime is generally a worse state of affairs than being an oppressed woman in a developed nation or a democratic polity. Committed to the task of dispelling misunderstandings and creating alliances between women in developed nations and women in developing nations in particular, global feminists aimed to widen the scope of feminist thought. They insisted that feminists needed to attend to women's political, economic, and social problems as seriously and passionately as they have attended to women's sexual and reproductive concerns.

As they began to grow in number, global feminists summoned women in all nations to work together. Specifically, they urged women to view their differences positively rather than negatively and to use them to secure genuine freedom and well-being for each other. "Sisterhood" does not require a woman to keep her counsel for fear of offending others. On the contrary, it requires her to speak her mind, confident that those to whom she speaks can appreciate her difference as much as she can appreciate theirs.

Ecofeminism

Like multicultural and global feminism, ecofeminism not only strives to show the connections among all forms of human oppression but also focuses on human beings' attempts to dominate the nonhuman world, or nature, which has been associated with women from time immemorial. According to ecofeminists, we are connected to each other but also to the nonhuman world: animal and vegetable, at least. Unfortunately, human beings do not always recognize how much their own well-being depends on nature's well-being. As a result, we overuse and misuse the world's natural resources, we pollute the air and water with toxic fumes and materials, and we stockpile weapons of mass destruction. We do these things, said ecofeminists, because we falsely believe that we are benefiting ourselves, but in fact, we are only harming ourselves and future generations. No one, including women, can expect to be free in a regime that systematically destroys nature. Without nature thriving, human beings cannot expect to survive, let alone thrive.

Third-Wave Feminism

Most recently, a third school of feminist thought, so-called third-wave feminism, is emerging. Third-wave feminists are more than willing to accommodate diversity and change. They are particularly eager to understand the ways in which gender oppression and other kinds of human oppression cocreate and comaintain each other. For third-wave feminists, difference is the way things are. Their world is a Heraclitean world (an ever-changing one, consisting of the conflict of opposites), not a Parmenidean world (of everlasting stable essences or permanent categories). Moreover, contradiction and conflict are expected and even welcomed by third-wave feminists. They believe that if feminism

is to develop, it must learn how to better handle conflict and contradiction.

As part of their study of interlocking forms of oppression, third-wave feminists engage in research and writing that attends to the lives and problems of specific groups of women. A typical third-wave feminist text will include articles about women who represent a wide variety of multicultural perspectives: Hispanic American, African American, Asian American, Native American, and so on. Indeed, one would be pressed to find an article authored by a third-wave feminist that is not heavily hyphenated.

The fact that the majority of the U.S. population is no longer White has profound implications for the development of third-wave feminists. Third-wave feminists wonder what it means to be an ethnic or racial "minority" in a city or state where one's own kind of people constitute the largest group of people. They also point out that the always contested categories of race and ethnicity are being further tested by phenomena such as increased intermarriage and people identifying themselves as "some other race" than the racial categories traditionally used on U.S. Census Bureau forms. Clearly, being a third-wave feminist in the 2000s, where a growing number of people *choose* their racial or ethnic classification, is different from being a feminist in the days when racial and ethnic identity were largely imposed by White hegemonic forces.

Conclusion

Interestingly, some schools of feminist thought have had more of a direct impact on society than others. To be sure, psychoanalytic, care-focused, existentialist, and postmodern feminists have helped women better understand themselves and their identities. They have also helped society in general understand that gender—the concept of male/masculine and female/feminine—is per se contestable and that care is of equal value as justice. But it is the other schools of feminist thought that have so far led to concrete social change.

Liberal feminists have pushed women into the public realm, both in the political arena and in the paid workforce. They have brought to society's attention the need for women's suffrage, education, action, and advancement, and they have not been afraid of using affirmative action policies to strengthen women's social position.

Marxist-socialist feminists have put on society's collective radar socioeconomic issues such as the wage gap between men and women, equal pay and comparable worth controversies, and the devaluation of women's domestic work, especially their care work. Like care-focused feminists, Marxist and socialist feminists have underscored how wrongheaded it is to think of women's work as "free" work. Just because women care for the vulnerable family members (children, the infirm, the disabled, the elderly) at home and just because women cook and clean in their own houses, does not mean that their work does not contribute to the economy. If women were to refuse to do their "homework," then someone (men?) or some institution (the state?) would have to pay professional cleaners and care workers a substantial amount of wages, if only the minimum wage.

Despite the provocative nature of their writings, radical feminists have made sexual variation less of a taboo in society, and they have gone a long way to normalize lesbian and gay relationships and to improve sexual relations between men and women. Thanks to radical feminists, pornography, sexual harassment, prostitution, and certainly domestic violence and rape have become phenomena of great concern in society. In addition, radical feminists have pushed society to reconsider the pros and benefits of reproduction-controlling technologies (contraception, sterilization, and abortion) and reproduction-assisting technologies (donor insemination, in vitro fertilization, surrogacy, egg donation, and egg freezing).

Reflecting on all schools of feminist thought, it seems clear that multicultural, global, and ecofeminists have very much increased society's awareness of oppression. Oppression does take a variety of forms—sexual, racial, ethnic, national, religious—and these varieties of oppression clearly reinforce each other. Becoming aware of the ways in which we make life better or worse for each other is the first step for people of good will to take in the quest for social justice. Similarly, ecofeminists have been leaders in the animal rights, environmental improvement, and sustainability movements. Nature needs to be respected by culture, or the human community will find its future generations without water, fuel, and food.

Finally, third-wave feminists are shaping a new kind of feminism that is not so much interested in getting women to want what women *should* want as in responding to what women say they want and not second-guessing or judging whether their wants are authentic or inauthentic. Third-wave feminists describe the context in which they advocate such a feminism as one of "lived messiness." According to Heywood and Drake, part of this messiness includes

people who change their sex and/or gender and wear their multiraciality proudly. Third-wave feminists are pushing society's envelope, exploring the implications of living in a diverse society where the old categories of "us" and "them" are ambiguous and always changing.

Although feminists continue to have women and gender as their primary focus, most schools of feminist thought now stress that all pernicious isms—sexism, racism, classism, colonialism, nationalism, speciesism—need to be overcome if the status of women and men in all countries is to be equalized. Feminists will not feel their work is done until and unless all women live in societies where everyone's freedom and best interests are recognized and realized. There is as much a need for feminism today as ever before.

Rosemarie Tong

See also Essentialism; Feminist Critiques of Social Science Applications; Feminist Epistemology; Multiculturalism; Prejudice and Stereotyping; Oppression; Power; Sexuality; Situational Analysis

Further Readings

Beasley, C. (1999). *What is feminism? An introduction to feminist theory.* Thousand Oaks, CA: Sage.

Bunch, C. (1993). Prospects for global feminism. In A. M. Jaggar & P. S. Rothenberg (Eds.), *Feminist frameworks* (3rd ed., p. 251). New York, NY: McGraw-Hill.

Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity.* New York, NY: Routledge.

Cixous, H., & Clement, C. (1986). Sorties. In B. Wing (Trans.), *The newly born woman* (pp. 63–134). Minneapolis: University of Minnesota Press.

de Beauvoir, S. (2011). *The second sex* (C. Borde & S. Malovany-Chevallier, Trans.). New York, NY: Random House.

Echols, A. (1983). The new feminism of yin and yang. In A. Snitow, C. Stansell, & S. Thompson (Eds.), *Powers of desire: The politics of sexuality* (pp. 439–459). New York, NY: Monthly Review Press.

Firestone, S. (1970). *The dialectic of sex*. New York, NY: Bantam Books.

Heywood, L., & Drake, J. (1997). Introduction. In L. Heywood & J. Drake (Eds.), *Third wave agenda: Being feminist, doing feminism* (pp. 1–24). Minneapolis: University of Minnesota Press.

hooks, b. (1990). *Yearning: Race, gender, and cultural politics.* Boston, MA: South End Press.

Jaggar, A. M. (1983). *Feminist politics and human nature.* Totowa, NJ: Rowman & Allenheld.

King, Y. (1995). Healing the wounds: Feminism, ecology, and nature/culture dualism. In N. Tuana & R. Tong (Eds.), *Feminism and philosophy* (pp. 353–373). Boulder, CO: Westview Press.

Kittay, E. F. (1999). *Love's labor: Essays on women, equality, and dependency.* New York, NY: Routledge.

Lorde, A. (1995). Age, race, class, and sex: Women redefining difference. In M. S. Andersen & P. H. Collins (Eds.), *Race, class, and gender* (2nd ed., pp. 187–194). Belmont, CA: Wadsworth.

Lugones, M., & Spelman, E. (1999). Have we got a theory for you! Feminist theory, cultural imperialism and the demand for "the woman's voice." In J. A. Kourany, J. P. Sterba, & R. Tong (Eds.), *Feminist philosophies* (2nd ed., pp. 474–486). Upper Saddle River, NJ: Prentice Hall.

Mill, J. S. (1970). The subjection of women. In A. S. Rossi (Ed.), *Essays on sex equality* (pp. 123–243). Chicago, IL: University of Chicago Press.

Mitchell, J. (1971). *Women's estate.* New York, NY: Pantheon Books.

Nussbaum, M. C. (1999). *Sex and social justice.* New York, NY: Oxford University Press.

Rich, A. (1994). Compulsory heterosexuality and lesbian existence. In A. M. Jaggar (Ed.), *Living with contradictions: Controversies in feminist social ethics* (pp. 487–490). Boulder, CO: Westview Press.

Tong, R. (2008). *Feminist thought: A more comprehensive introduction* (3rd ed.). Boulder, CO: Westview Press.

Warren, K. J. (1997). *Ecofeminism: Women, culture, nature.* Bloomington: Indiana University Press.

Wollstonecraft, M. (1975). *A vindication of the rights of woman.* New York, NY: W. W. Norton.

Young, I. M. (1981). Beyond the unhappy marriage: A critique of the dual systems theory. In L. Sargent (Ed.), *Women and revolution: A discussion of the unhappy marriage of Marxism and feminism* (pp. 4–6). Boston, MA: South End Press.

Feminist Critiques of Social Science Applications

This entry assesses the relationship between feminist research and social-scientific analysis. Although employing many of the tools of social-scientific research, feminists have been critical of those tools and developed uniquely feminist approaches to the social sciences.

From the beginning of the contemporary feminist movement, feminists have had an uneasy relationship to the social sciences. On the one hand, some of the core assertions of contemporary feminism are

rooted in the social sciences. Beginning with Simone de Beauvoir's *The Second Sex,* feminists have made empirical claims about the status of women in society: their inferior position, the discrimination against them, and the denigration of the feminine in our culture. These social-scientific facts have formed the core of contemporary feminism. On the other hand, however, feminists have also claimed that the social sciences are inherently masculinist and that this bias pervades all social-scientific disciplines. They assert that this masculinist bias marginalizes the experience of women in society and precludes an accurate assessment of the status of women.

Perhaps the most contentious issue in the relationship between feminism and the social sciences is the status of empiricism. At the root of much, although not all, social-scientific analysis is the set of assumptions about facts, values, knowledge, and objectivity that defines empiricism. From the outset, feminists have been critical of empirical social science, claiming that it renders the experiences of women invisible. They further claim that the objectivity that is the goal of empirical research is illusory. This has led many feminists to reject empiricism out of hand and adopt alternative approaches to social-scientific analysis. But this rejection is neither universal nor unambiguous. A total rejection of empiricism has consequences that put feminists in an awkward position. If feminists cannot assert the truth and objectivity of claims about the status of women in society, then much of the force of feminist critique is lost.

As a way out of this dilemma, some feminists have called for a "new empiricism" to guide feminist research in the social sciences. Feminists such as Sandra Harding and Lynn Hankinson Nelson have argued for a definition of empiricism that recognizes that all knowledge is situated and perspectival. While not rejecting the notion of objective knowledge that grounds empiricism entirely, these feminists argue that we must significantly redefine it in a way that takes account of this situatedness and the reality of women's lives. They also claim that all knowledge is biased and that objective, value-free knowledge is unobtainable. These theorists agree with Dale Spender's claim that there is no one truth, no one authority, and no one objective method that leads to the production of knowledge. But they also claim that we can and do seek knowledge of women's lives in society and that this knowledge is valid.

Feminists' skepticism about empiricism has caused some to turn to alternative approaches to the social sciences. But in adopting these approaches, feminists have in every case developed a method that is critical and distinctively feminist. One approach is phenomenology. Feminists such as Dorothy Smith assert that feminist research must begin from an examination of the lived experience of women. Although phenomenology emphasizes lived experience, phenomenologists have, to a large extent, ignored women's experiences. Feminist phenomenologists have sought to correct this by focusing on the actuality of women's lives.

A closely related feminist approach has its roots in Marxism. Marxist standpoint theory offers an alternative to the objectivism of empirical social science. But for Marxists, standpoint theory was only applied to the position of the proletariat; women did not figure in the equation. Feminist standpoint theory, developed by theorists such as Nancy Hartsock, extended Marx's theory to the sphere of women. Arguing that women and men have different standpoints and hence different conceptions of knowledge, Hartsock and other standpoint theorists argued that women, like the proletariat, have a clearer understanding of the truth of society because of their marginalized position.

In the latter half of the 20th century, many disciplines in the humanities and social sciences were caught up in what has been referred to as the *linguistic turn.* Adherents argued that language is the single constituent force of social reality and that social analysis is exclusively linguistic analysis. Feminists were deeply influenced by this movement. Linguistic analysis, most particularly postmodernism, became the focus of much of feminist analysis. Feminists' analyses of concepts such as "woman" and "sex" produced much valuable work, clarifying the origins of the inferiority of women in society. But again, the feminist approach to linguistic analysis was distinctive and oppositional. Women were missing from most linguistic analyses in the social sciences. Feminists, focusing on how language constructs the social reality women inhabit, sought to redress this omission.

In recent years, the dominance of the linguistic turn has been challenged by an approach labeled the "new materialism." The new materialists seek to bring the material back into the equation without losing the advantages of the linguistic turn. Feminists have been on the forefront of this movement with their analyses of the materiality of women's bodies.

Along with the new materialists in the natural and social sciences, feminists are reshaping our conception of truth and knowledge.

Feminists have been and will continue to be concerned with the status of women in society and how that status can be improved. This brings them in close connection to the social sciences. But feminism can never be simply classified as one of the social sciences. Feminists' insistence on perspectival knowledge, an oppositional stance, advocacy, and the critique of objectivity flies in the face of mainstream empirical social science. Feminists are and must remain critical, challenging the fundamental tenets of established social science.

Susan Hekman

See also Androcentrism and the Philosophy of Science; Empiricism; Feminism: Schools of Thought; Feminist Economics; Feminist Epistemology; Marxism and Social/Historical Explanation; Postmodernism

Further Readings

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning.* Durham, NC: Duke University Press.

Butler, J. (1990). *Gender trouble.* New York, NY: Routledge.

de Beauvoir, S. (2010). *The second sex* (C. Borde & S. Malovany-Chevalier, Trans.). New York, NY: Knopf.

Harding, S. (1991). *Whose science? Whose knowledge? Thinking from women's lives.* Ithaca, NY: Cornell University Press.

Hartsock, N. (1983). *Money, sex, and power.* New York, NY: Longman.

Nelson, L. H. (1990). *Who knows? From Quine to a feminist empiricism.* Philadelphia, PA: Temple University Press.

Smith, D. (1987). *The everyday world as problematic: A feminist sociology.* Boston, MA: Northeastern University Press.

Spender, D. (1980). *Man-made language.* London, England: Routledge.

Feminist Economics

Feminist critiques of economics emerged as a counterpoint to the economic theories built around core ideas of self-interested individualism and contractual exchange. The conceptual framework of 20th-century economics focused on the marketplace and relegated nonmarketed human interactions to its marginalized fringes or to other disciplines, largely ignoring economic life outside the market—arenas in which women's lives are disproportionately involved. With the establishment of the International Association for Feminist Economics and its journal, *Feminist Economics,* in the early 1990s, feminist economics emerged as a distinct area of economic inquiry, broadly focusing on reconstructing economic theory to present theories and policy tools more responsive to women and their lives and promoting of gender justice. Calling attention to theoretical constructions that deemphasized or ignored women's lives, such as the idea that only remunerated employment should be counted as productive work, feminist economists argued that the whole economy includes interdependent paid and unpaid sectors, each influenced by policies directed toward the other, and that economic policies should promote the flourishing of all human life rather than focusing exclusively on measures of economic output or growth. Feminist economics is growing rapidly and putting forward new insights into all areas of economics. This entry describes several of the most important ideas and their implications for economic policy.

Theorizing Gender Inequality

Amartya Sen calculated that in the 20th century more than 100 million women have died disproportionately to men because of lack of equal access to food, medical attention, and other survival-related resources. Subsequent studies replicated this result, revealing that more women and girls have died from causes relating to gender inequality than those who died in both world wars. With further evidence of pervasive gender inequality occurring in education, health care, work, wealth, and income, feminist scholars have argued that gender inequality should be treated as a central economic concern meriting theoretical explanation and policy remedies.

Gender Relations and the Family

In challenging economics to provide more complete theories, feminist economists have called for models that look within the family to examine how gender relations influence economic outcomes. Since economic approaches that rely on aggregated measures

of economic output, such as gross domestic product, do not reveal how resources are distributed within families, they obscure gender inequality and provide no help in illuminating either its causes or its impact.

Standard economic models treat the family either as an amicable homogeneous unit or as a civilized bargaining game. In contrast, feminist economists point to unequal gender relations, ranging from simple devaluations of women and their work to exploitative gender relations often accompanied by violence. Linking women's lower economic status to their greater role in unpaid reproductive and care work, feminist economists have sought to understand how social norms relating to such work influence economic distributions in the home and marketplace.

Human Interconnections and Care Work

Whereas standard economic models treat people as independent agents, able to operate in markets and take care of themselves, feminist economists point to the interconnectedness of human life. One aspect of human interconnectedness is the need for care and support from others, beginning at birth. Care work, disproportionately unpaid and performed by women, provides essential support not just for children but also for the sick, the disabled, and the elderly and, more generally, for the daily work of cooking, cleaning, and caring for families.

Feminist economists argue that work doesn't have to be paid to be valuable and that this insight has important implications for economics. For example, if economic models do not acknowledge the value of care work, policies based on such models may have unanticipated consequences and adverse outcomes. Cutbacks in health care may mean that sick people will remain ill longer and therefore need more care from family members. The caregiver, typically a woman, will need to reduce paid work hours or work more intensively; or those who need care may not get enough of it, resulting in greater morbidity and a higher likelihood of death. Cutbacks in publicly provided child care may similarly lead to more demands on family caregivers; or children may just receive less care and have fewer skills, reduced productivities, and lower well-being. Feminist economists argue that such examples reveal the benefit of economic approaches that recognize the interdependence between the marketed and unmarketed sectors of the economy. Feminist research also provides insight into other aspects of family life and domestic relationships, including variations resulting from other family configurations, such as same-sex and single-mother households.

Gender-Aware Economic Policies

A widely held view among feminist economists is that gender justice should be a goal of economic policy and that efforts to end gender inequality must be based on gender-aware economic models that more fully reflect all human lives and not just those of male or European-descended populations. Gender mainstreaming, the idea that public actions should be thoroughly evaluated and monitored to provide equal benefit to women and men, is one such economic policy. Gender budgets similarly are designed to ensure that public expenditures have equal impact on women and girls as compared with men and boys. Feminist economic research, more broadly, points to the potential success of a wide variety of policies designed to enhance gender justice and women's well-being.

Diana Strassmann

See also Androcentrism and the Philosophy of Science; Feminist Critiques of Social Science Applications; Feminist Epistemology; Heterodox Economics; Markets and Economic Theory; Prejudice and Stereotyping

Further Readings

Agarwal, B. (1997). "Bargaining" and gender relations: Within and beyond the household. *Feminist Economics, 3*(1), 1–51.

Benería, L. (2008). The crisis of care, international migration, and public policy. *Feminist Economics, 14*(3), 1–21.

Benería, L., May, A. M., & Strassmann, D. L. (Eds.). (2011). *Feminist economics* (3 vols.). Northampton, England: Edward Elgar.

Klasen, S., & Wink, C. (2003). "Missing women": Revisiting the debate. *Feminist Economics, 9*(2–3), 263–299.

Strassmann, D. L. (1994). Feminist thought and economics; or, what do the Visigoths know? *American Economic Review, 84*(2), 153–158.

Feminist Epistemology

This entry presents, both historically and thematically, the birth and development of feminist epistemology.

Why Feminist Epistemology?

Succinctly put, the mission of feminist epistemology is to supply an answer to the question "Is the sex of the knower epistemologically significant?"—which was posed by Lorraine Code in a 1981 article by that title. The question, which is germane to postpositivist knowledge-producing projects in the physical and social sciences and in everyday life, does not admit of an easy "yes" or "no" response. In a tradition where the goal of achieving objective certainty in knowledge depended on the interchangeability of knowers in replicable observational conditions, the idea that any subjective property—such as the sex—of such knowers could claim epistemological significance contests the basic principles of empirical inquiry. Although the question may sound facile, its implications are complex and far-reaching. It asks whether certain specificities of subjectivity could count among the conditions that make knowledge possible or could impede its achievement. It comes neither from a simplistic thought that what is true and valid for women might not be true and valid for men, and vice versa, nor from an urge to promote "women's ways of knowing," but from a conviction that the diverse circumstances of human lives from which subjectivities are constituted are such as to contest the reductivism from which universal human sameness—and epistemic interchangeability—are presupposed.

For non-philosophers, epistemology is an esoteric inquiry, relevant to philosophers alone; but reliable, authoritative knowledge is integral to feminist practices of developing informed analyses of social-political oppression and marginalization and engaging in emancipatory projects. In addition, epistemologists seek to articulate principles and procedures for evaluating knowledge claims for their veracity and their social-political implications. Feminists have to *know* the social-structural implications of women's marginalization in public knowledge, as knowers and known, so as to account for women's continued underrepresentation—their minimal epistemic power and authority—across intersections of gender, race, class, ethnicity, sexuality, ability, and disability. Hence, they have to know the social, discursive structures well enough to show how they sustain hierarchies of power and privilege, in projects that frequently involve learning to see what was previously invisible, taken for granted, or dismissed as showing that women are incapable of the "best" kinds of knowledge. Thus, while "mainstream" epistemologists evaluate sources of evidence and methods of inquiry that are generically conceived, seeking universal, necessary and sufficient conditions for achieving certainty and refuting skepticism, feminist epistemologists are also committed to analyzing the nature and positioning of knowers and the (gendered) politics of knowledge and to discerning what "universal" conclusions fail to address.

Historically, Western philosophy has aligned mind/reason with maleness and body/emotion with femaleness, representing emotions as irrational and invoking women's alleged emotionality to disqualify them as knowers. Yet feminists, indebted to Genevieve Lloyd's *The Man of Reason,* have shown that maleness or femaleness are not incidental to but constitutive of subjectivities and cognitive possibilities, variably throughout Western history. In consequence of "different" (from affluent White male) embodiment, people have been assigned marginal epistemic status by criteria that assimilate knowing to a single norm, and they thence assume universal access to uniform experiences for which differences are dismissed as irrelevant aberrations yet exposed as exclusionary by feminists and other Others.

Informed by the consciousness-raising practices of the 1960s, feminists, latterly in concert with antiracist and other postcolonial theorists, have exposed cognitive dissonances and incongruities between women's multiple and diverse experiences and theories and conceptual schemes implicitly but routinely derived from a White male norm, which purport to explain—to *know*—the physical–social–material world and to know women's and other Others' experiences as a part of it. Eschewing aims of determining how ideal, abstract, and infinitely replicable knowers ought to know, feminists such as Donna Haraway in *Situated Knowledges* and Lorraine Code in *What Can She Know?* have produced sophisticated, critical-constructive analyses of how diversely situated and embodied knowers can

and do know and how normative epistemic requirements can and should be fulfilled by real knowers.

Beginnings

In the 1980s and early 1990s, feminist epistemologists tended to occupy one of two positions: (1) feminist empiricism or (2) feminist standpoint theory. Some, such as Sandra Harding in *The Science Question in Feminism* and Helen Longino in *Science as Social Knowledge,* accorded biology and the social sciences the centrality, if not precisely the paradigm status, post-positivist epistemology had accorded the physical sciences. Others, including contributors to Linda Alcoff and Elizabeth Potter's *Feminist Epistemologies,* drew on mainstream epistemology as well as its philosophy-of-science offspring. In the 21st century, feminist epistemology can no longer adequately be mapped within these categories, even by enlisting postmodernism—the third initial category to address the remainder; yet the postmodern import of the *antiessentialism* that characterizes all three positions is apparent.

Briefly, feminist empiricists focus on evidence gathering, advocating a method alert to the exclusionary effects of androcentrism in judgments of what counts as evidence worthy of note, while standpoint theorists focus on the constitutive effects of the historical-material positioning of women's lives, practices, and experiences. For feminists, an empiricism committed to evidence gathering and justification that honors objectivity, yet is informed by feminist insights and commitments, can produce more adequate knowledge than classical empiricism, oblivious of its complicity in sustaining a pervasive sex/gender system. An enhanced vision enables feminists to enlist empiricist tools to expose the sexism, racism, and other "isms" that, often silently, inform knowledge production and circulation. In Longino's social empiricism, communities, not individuals, are knowers: Their background assumptions contribute to shaping knowledge as process and product. Taking genetic research as illustrative, Longino shows how assumption (value)-driven differences in knowledge production contest the possibility of value neutrality while endorsing respect for evidence and collaborative cognitive agency. Lynn Hankinson Nelson, in *Who Knows: From Quine to a Feminist Empiricism,* develops from "naturalized epistemology" a neo-empiricism for which communities, not individuals, are the primary knowers; and knowers come to evidence through webs of belief, available for communal critique. Because the marginalized cannot realize their emancipatory goals without understanding the intractable aspects of the physical and social world and its malleable, contestable features, they have to achieve a fit between knowledge and "reality," even when "reality" consists of social artifacts such as racism, oppression, or pay equity. For feminist empiricists, an empiricism alert to gender specificity in its multiple intersectional modalities can achieve just this. Thus, for Sandra Harding, politically informed inquiry introduces a "strong objectivity," to yield a better empiricism than inquiry conducted from an allegedly neutral "view from nowhere" can achieve.

For standpoint theorists such as Nancy Hartsock, in *Money, Sex, and Power: Toward a Feminist Historical Materialism,* empiricists lack the conceptual resources to address the historical-material diversity from which people produce knowledge. Standard-setting knowledge in Western societies has derived from the experiences of White, middle-class, educated men, with women (like Marx's proletariat) occupying underclass epistemic positions. As capitalism "naturalizes" the subordination of the proletariat, patriarchy "naturalizes" the subordination of women, and as engaging with material-social experiences from the standpoint of the proletariat denaturalizes these assumptions, so does starting from women's lives denaturalize the patriarchal order. A feminist standpoint is a hard-won product of consciousness raising and social-political engagement, in which the knowledge that enables the oppressed to survive under oppression is enlisted as a resource for initiating social transformation.

Neither empiricist nor standpoint feminism can resolve all the issues. Even the new empiricism falls short of fully addressing the power-saturated circumstances of diversely located knowers or of engaging with questions about how evidence is discursively constituted and whose evidence it suppresses in the process. Nor can standpoint theory alone avoid obliterating differences. Its "locatedness" offers a version of social reality as specific as any other, albeit distinguished by its awareness of that specificity. But empiricism's commitment to exposing the concealed effects of gendered and other specificities in knowledge production cannot be gainsaid, nor can standpoint theory's critical analyses of women's experiences, with its focus on how hegemonic values, entrenched power structures, and

taken-for-granted assumptions about human sameness work to perpetuate marginalization.

In the decades since empiricism and standpoint theory seemed to cover the territory, with postmodernism posing anti-epistemological challenges to both, feminists have found these alternatives neither mutually exclusive nor able, separately or together, to adequately explain the sexual politics of knowledge production and circulation. Cross-fertilizations have proven more effective than adherence to methodological orthodoxy: The idea of a unified theory, or of the unity of knowledge, has—productively—been discredited in its putative status as a regulative ideal.

Differences

Feminists cognizant of the differences that difference makes do not, in fact, hope to achieve a unified standpoint. Thus, Patricia Hill Collins, in *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment,* advocated an "outsider-within" Black feminist standpoint: an Afro-centered epistemology that shows how knowledge produced in racially subordinated groups fosters resistance to the hegemonic knowledge produced at the putative "center." And Maria Lugones, in "Playfulness, 'World-Travelling' and Loving Perception," advocated practices that eschew, and move self-consciously away from, a too particular, self-satisfied location at a hitherto uncontested social and epistemic center, to travel, literally or metaphorically, to "other worlds" and learn how it is there. Donna Haraway recast the subject and object of knowledge as radically located and unpredictable and knowledge construction as learning to see from positions discredited in dominant accounts of knowledge and reality. Taking women's cognitive practices seriously can enable feminists to eschew the individualism and universalism of mainstream theory, to examine specifically situated knowing, where theory, materiality, and practice are reciprocally constitutive and where knowers are diversely positioned and active within them. Thus, Evelyn Fox Keller's biography of the geneticist Barbara McClintock, *A Feeling for the Organism,* presents a scientist engaged with her objects of study and attuned to differences, as evidence for a psychosocially gender-inflected style of research. And in *What Can She Know?* and *Ecological Thinking,* Lorraine Code examines how power and privilege yield asymmetrically gendered, racialized, and class-inflected standards of authority in medical knowledge, in the experiences of welfare recipients, for testimonial credibility, and in women's responses to sexist and racist challenges. Her ecological model of knowledge and subjectivity unsettles the "master" model that governs mainstream epistemology.

Recent Developments

Noteworthy for this account are developments that build on and incorporate aspects of the early positions while no longer aligning with the original taxonomy. The issues to be discussed in this section are the significance of testimony for social epistemology; the blurring of boundaries between epistemology, ethics, and political theory; and the new epistemologies of ignorance.

For Anglo-American epistemologists, the principal sources of empirical knowledge were perception, memory, and testimony, listed in this order to show their comparative levels of reliability—perception being the most reliable because it is the most "grounded," while testimony, with its reliance on other people, ranking a distant third, was often assimilated to reliance on opinion or hearsay rather than on the putative security of direct perception and memory. Recent developments in social epistemology, to which feminists are among the most innovative contributors, contest this ordering, returning epistemology to the real world, where it is clear that only a very small proportion of what people reasonably claim to know is theirs by virtue of direct perceptual experience or memories thereof. "We," whoever we are, are multiply interdependent in our knowledge-productive activities: Whereas formal analytic epistemology sought to transcend interdependence, social epistemologists—and feminists prominently among them—take it as a given and analyze its multiple modalities. Clearly, reliance on testimony as a source of knowledge requires taking the subjectivities of testifiers into account, for decisions about who can be trusted become crucial. It is in engaging with this issue that some of the most productive recent feminist inquiry has been produced. For theorists who engage with the centrality of testimony, questions such as those Lorraine Code raises in *Epistemic Responsibility* and Miranda Fricker in *Epistemic Injustice* and issues in the politics of incredulity move to central places in epistemological inquiry. The tasks facing epistemologists become increasingly complex as empirical observational inquiry is itself shown always to be inflected

by the social-political positioning of knowers and the known, and conceptions of objectivity (which is by no means eschewed in such inquiry) become entangled with the assumptions born out of a dominant social-epistemic imaginary whose constitutive part in shaping received views of knowledge and epistemic subjectivity has itself to be subjected to analysis.

Crucial too in understanding the diverse facets of these developments are the new epistemologies of ignorance discussed by the contributors to Shannon Sullivan and Nancy Tuana's edited volume *Race and Epistemologies of Ignorance*, where theorists examine the moral-political effects of the ignorance inadvertently or otherwise produced by, or alongside, received bodies of knowledge. All these developments show how feminist epistemology continues to be a vibrant and challenging field of inquiry. Some of these thoughts come to fruition in texts such as George Yancy's collection *The Center Must Not Hold*, whose title and content capture a governing insight and inspiration of the best feminist and anti-racist epistemology of the early 21st century.

Lorraine Code

See also Androcentrism and the Philosophy of Science; Embodied Cognition; Epistemology; Essentialism; Feminism: Schools of Thought; Feminist Critiques of Social Science Applications; Feminist Economics; Kinds: Natural Kinds Versus Human Kinds; Postcolonial Studies; Postmodernism; Situated Cognition; Situational Analysis; Social Epistemology; Trust, Epistemic; Virtue Epistemology

Further Readings

Alcoff, L., & Potter, E. (Eds.). (1993). *Feminist epistemologies*. New York, NY: Routledge.

Code, L. (1991). *What can she know? Feminist theory and the construction of knowledge*. Ithaca, NY: Cornell University Press.

Code, L. (1995). *Rhetorical spaces: Essays on (gendered) locations*. New York, NY: Routledge.

Code, L. (2006). *Ecological thinking: The politics of epistemic location*. New York, NY: Oxford University Press.

Collins, P. H. (2000). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment* (2nd ed.). London, England: Routledge.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford, England: Oxford University Press.

Grasswick, H. E. (2011). *Feminist epistemology and philosophy of science: Power in knowledge*. Dordrecht, Netherlands: Springer.

Haraway, D. (1990). Situated knowledges. In D. J. Haraway (Ed.), *Simians, cyborgs and women* (pp. 183–202). New York, NY: Routledge.

Harding, S. (1986). *The science question in feminism*. Ithaca, NY: Cornell University Press.

Harding, S. (2008). *Sciences from below: Feminisms, postcolonialities, and modernities*. Durham, NC: Duke University Press.

Hartsock, N. (1983). *Money, sex, and power: Toward a feminist historical materialism*. New York, NY: Longman.

Keller, E. F. (1983). *A feeling for the organism: The life and work of Barbara McClintock*. New York, NY: W. H. Freeman.

Lloyd, G. (1984). *The man of reason: "Male" and "female" in Western philosophy*. London, England: Routledge.

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton, NJ: Princeton University Press.

Lugones, M. (1989). Playfulness, "world-travelling" and loving perception. In A. Garry & M. Pearsall (Eds.), *Women, knowledge and reality* (pp. 419–434). Boston, MA: Unwin Hyman.

Nelson, L. H. (1990). *Who knows: From Quine to a feminist empiricism*. Philadelphia, PA: Temple University Press.

Rose, H. (1983). Hand, brain and heart. *Signs, 9*, 73–90.

Scheman, N. (2011). *Shifting ground: Knowledge and reality, transgression and trustworthiness*. New York, NY: Oxford University Press.

Shotwell, A. (2011). *Knowing otherwise: Race, gender, and implicit understanding*. University Park: Pennsylvania State University Press.

Sullivan, S., & Tuana, N. (Eds.). (2007). *Race and epistemologies of ignorance*. Albany: SUNY Press.

Yancy, G. (2010). *The center must not hold: White women philosophers on the whiteness of philosophy*. Plymouth, England: Lexington.

Feyerabend, Critique of Rationality in Science

This entry introduces the work of Paul Feyerabend and reviews in particular his criticisms of scientific rationality as well as counterarguments and rejoinders. Feyerabend has been an influential critic of

accounts of scientific rationality and the role scientific rationality plays in our lives. His criticisms of scientific rationality are different in different periods of his career. However, there is an important underlying assumption throughout his work, which is that the choice of a scientific way of life is primarily an ethical choice. It is not one dictated by rational or empirical canons of rationality.

In the work Feyerabend wrote between the 1950s and the early 1970s, his criticisms focus on naive empiricist pictures of science and on narrow conceptions of the scientific enterprise. Like Karl Popper, he sees science properly conducted as part of a critical attitude that has an important and positive role in human life. The 1961 lectures titled *Knowledge Without Foundations* contain a clear encapsulation of his early view. In them, he argues that the scientific attitude involves two components: first, producing bold and broad theories that go beyond and against appearances and, second, producing and seriously considering important criticisms of those theories.

It is Feyerabend's later work that has been highly influential. In that work, his criticisms focus on the role of science in society. He sees that role as rationally unjustifiable and ethically dubious.

Critique of Scientific Rationality

Feyerabend's later work contains at least four criticisms of scientific rationality:

1. The first criticism already occurs in his early work. It is a criticism of philosophical reconstructions of scientific reason and not of science. The criticism is that there is no universal scientific method in the form of fairly precise formal rules that has been used by successful scientists or by the scientific community. This criticism was highlighted in the book *Against Method,* which first appeared in 1975. In it, Feyerabend defends the principle, ironically, that if one wants to make progress in science, anything goes. He calls this the principle of "epistemological anarchism." His primary argument for his claim is contained in a careful case study of Galileo's fundamental contribution to the Copernican revolution.

Feyerabend argues that Galileo proceeds counterinductively rather than inductively in producing his version of Copernicanism; relies on poor observations, which he rhetorically puffs up; defends inertia in an ad hoc manner; and bolsters his theory through various forms of trickery. Furthermore, many of Galileo's predictions on the basis of his version of Copernican theory were falsified.

Feyerabend's friend Imre Lakatos was inclined to agree with much of what Feyerabend had argued. However, he attempted to create a framework for assessing the rationality of research programs in the long run, which would have allowed Galileo and others to get off the hook. On Lakatos's account, it is essential to science for some individual scientists to play a risky game by producing bold hypotheses that go against the observable evidence. Sometimes such strategies pay off because the bold hypotheses eventually predict great numbers of startling novel facts and radically simplify and reorganize the known information. Nevertheless, the scientific community as a whole is rational because journal editors and funding bodies will eventually (rightly) prefer research programs that predict novel facts and unify knowledge to those that have failed to do so.

In response, Feyerabend argues in *Against Method* that Lakatos's theory is a form of epistemological anarchism in disguise—Lakatos has failed to show that any move by scientists or the scientific community could be rationally ruled out. After all, Lakatos admits that some scientific theories, like atomism, had only paid off after thousands of years and had been permitted over a long time to continue to exist even though they had not predicted a single novel fact for ages. All Lakatos does is call any move by scientists that happened to be regarded as successful later "rational."

Feyerabend does not seem to have intended his first criticism to undermine the status of science. Rather, he seems to have intended to agree with Thomas Kuhn that scientific method is learned through examples of good scientific work and is not reducible to general rules.

Recent commentators have been inclined to agree with Feyerabend's criticism of simplistic formal accounts of a universal method. However, his analysis of the Galileo case has been undermined by detailed historical work. This work has exposed a number of important problems, including the following. Galileo did not introduce the notion of inertia ad hoc. Rather, he introduced it after careful experiments with inclined planes. His defense of the use of the telescope involved careful work that showed in detail that ordinary observation of bright objects at night is worse than telescopic observation. Furthermore, Galileo used the commonsense

principle of inference to the best explanation to justify many of his claims.

2. Feyerabend's second criticism of scientific rationality relies on his claim that science contains incommensurable theories. On Feyerabend's account, two theories are incommensurable if they are comprehensive cosmological schemes, and if not, a single primitive descriptive term of one theory can be defined in terms of any primitive descriptive term of the other. He claims that Newtonian mechanics and Einstein's Relativity theories are incommensurable in this way. Both deal with the entire universe and everything in it. Each describes any item in a way that presupposes a radically different view about the nature of things. For instance, the mass of an object is an intrinsic property in Newtonian mechanics—namely, the amount of matter in an object. In Einstein, it is a relational quantity. This means that not only this property but also what it is to be an object vary between the two theoretical frameworks.

Incommensurable theories cannot be tested using descriptions that are independent of the theories, as the theories are comprehensive. For example, we cannot test Einstein's claims about mass while using the term *mass* in its Newtonian sense, since our description of the experimental test presupposes that Newtonian mechanics is true. In his early work, Feyerabend argues that this is not a serious problem unless we are wedded to some naive empiricist ideas about scientific testing. He proposes methods for testing the relative merits of incommensurable theories. Each theory can be tested by using its own descriptive terms to describe the result of an experiment. If there is an inconsistency between the predicted result and the result obtained, then the theory has been refuted using its own descriptive terms.

By the time he wrote *Against Method* and his more radical 1978 work, *Science in a Free Society,* Feyerabend had come to think that incommensurability posed much deeper problems for the rationality of science. The reason is threefold: First, the ontologies of two comprehensive cosmological schemes sometimes differ far more radically than he had initially thought. Second, the most fundamental standards for judging truth sometimes differ between theories. Third, comprehensive cosmological schemes constitute the world—they do not merely describe an independently existing world.

Feyerabend spells out his claims primarily through a detailed discussion of the worldview of the archaic Greeks shortly before the appearance of philosophy. He claims that Homer treats gods as everyday observables that we would describe instead as psychological states. For instance, what is perceived by Homer's heroes as a direct experience of Athena is, we would say, a sudden rush of adrenaline. Dreams are perceived by Homer's heroes as things that are sent by the gods and literally arrive top down. They are not mental events. The conception of knowledge in Homer is radically different—knowledge is conceived of as lists not as theories. Feyerabend claims that this makes independent tests impossible. Indeed, he claims that the disappearance of the Homeric worldview at the hands of Greek philosophers was literally the disappearance of gods and their replacement by things like atoms in empty space.

Feyerabend's second criticism seems to collapse into incoherence, for if he is right, one cannot be thinking about the same things when one is thinking as a Homeric Greek and when one is thinking as a modern westerner. This implies that Feyerabend cannot be explaining to us in our world what people in the Homeric world are thinking. This is because (a) if Homeric people are in the other world, they are not in our world, and if they are in our world, they are not in the other world and (b) Homeric descriptions will not refer to anything in our world. What we understand to be Homeric talk in our world thus cannot be incommensurable with our talk, and Feyerabend cannot prove his case that there is another, radically different world.

Feyerabend was inclined to water down incommensurability to deal with this objection. In later works, he criticizes relativism, though he still argues that when dealing with very different outlooks, simple-minded comparisons are trivial or irrelevant. He seems to continue to argue that we cannot reasonably judge that a very different outlook is false or inferior from a transcendent standpoint. However, critics of Feyerabend would respond that this is irrelevant if outlooks are not incommensurable with one another.

3. Feyerabend's third criticism of the rationality of science is aimed not so much at science but at the role science plays in our lives. He argues that whether science tells us the truth or not, it should not have to play a crucial role in our lives. Whether scientific knowledge is useful or not depends on how we choose to live. We might rationally decide that a way of life in which science plays very little part is better than a way of life in which it does. Science

may well reveal a world that is dull, offers us little consolation, is threatening, and so on. We might well be better off living with consoling falsehoods than with unpleasant truths. Furthermore, science does not exist in abstraction from real societies. On the contrary, it is developed in real societies that have structural tendencies to apply it in ways that threaten human interests. Feyerabend criticized the demand for logical preciseness in a similar way when he noted that "a logically precise idea of tap dancing might give us persistent cramps." Logic, like science, not only exists as a bunch of abstractions, it is used by real people in real situations, and its effects on real people need to be studied before we can confidently pronounce that its widespread use will be useful. A useful illustration of this point can be found in the documentary film *The Fog of War.* In the film, the former U.S. Secretary of Defense Robert McNamara praises his training in logic. He then goes on to explain that in the U.S. Air Force in World War II he used the skills he learned to help him calculate the maximum damage that could be achieved from minimum input in firebombing Japan.

As a consequence of this third criticism, Feyerabend argues that we should separate the state from science in the same way in which advanced societies had separated the state from religion. Citizens should be free to choose what role science plays in their lives and the lives of their children.

Arguments against the third criticism include the following. First, even those who reject a scientific way of life will need to know a considerable amount about science to protect themselves from its effects. A romantic primitivist will need to know something about the effects of nuclear power to deal with a proposal to build a nuclear power station. Second, the use of nonscientific standards by citizens often constitutes a serious threat to others. For instance, medieval views about witches and how to discover them pose a serious threat to those accused of witchcraft. It is legitimate to prevent such attitudes of mind from developing by educating people to judge matters scientifically. Third, children need to be protected from the actions of their parents. Most people do not find science threatening. Indeed, they find science very useful. To deprive children of a scientific education because of the views of their parents is unjustified. It should be noted that none of these arguments is a decisive and complete rebuttal of Feyerabend's third criticism of the rationality of science. Feyerabend thinks that such arguments at best show that the separation of the state from science needs to be qualified.

4. Feyerabend's fourth criticism of the rationality of science is related to the third criticism. It is best put in his posthumously published work *The Conquest of Abundance*. He argues that science creates abstractions and idealizations, which, though useful in limited ways, fail to capture the richness and diversity of being. Science is of some limited use to us, but this gives it no rational warrant to replace other aspects of our picture of the world. This criticism is tied to the claim that various nonscientific ways of conceptualizing and dealing with being can be successful according to a range of criteria for success. Some of his central arguments for this criticism cannot be easily explained abstractly as they are analyses of episodes from the history of philosophy, of science, and of cross-cultural contacts. A central theme in those analyses is that philosophical and scientific arguments for views about being presuppose the merits of rationalism or science—they do not show that science and rationalism are correct. There has not been much debate about the fourth criticism. Nevertheless, it may be argued that we do not need rationalism and science to judge science as being superior to traditional views in important respects. We can instead use commonsense criteria to choose predictively powerful and more coherent theories. Science is on a continuum with some parts of everyday thought in every human culture. This, of course, does not indicate that rationalism and science are superior in every respect to other belief systems. So perhaps Feyerabend is partly right.

Whether parts of Feyerabend's critique of scientific rationality are cogent is still a matter of considerable debate.

George Couvalis

See also Kuhn on Scientific Revolutions and Incommensurability; Lakatos, Methodology of Scientific Research Programs; Observation and Theory-Ladenness; Popper's Philosophy of Science; Relativism in Scientific Theories; Science and Ideology

Further Readings

Couvalis, G. (1989). *Feyerabend's critique of foundationalism.* Aldershot, England: Avebury.

Couvalis, G. (1997). *The philosophy of science: Science and objectivity.* London, England: Sage.

Farell, R. (2003). *Feyerabend and scientific values: Tightrope-walking rationality.* Dordrecht, Netherlands: Kluwer.
Feyerabend, P. (1978). *Science in a free society.* London, England: New Left Books.
Feyerabend, P. (1981a). *Philosophical papers: Vol. 1. Realism, rationalism and scientific method.* Cambridge, England: Cambridge University Press.
Feyerabend, P. (1981b). *Philosophical papers: Vol. 2. Problems of empiricism.* Cambridge, England: Cambridge University Press.
Feyerabend, P. (1987). *Farewell to reason.* London, England: Verso.
Feyerabend, P. (1991). *Three dialogues on knowledge.* Oxford, England: Blackwell.
Feyerabend, P. (1995). *Killing time: The autobiography of Paul Feyerabend.* Chicago, IL: University of Chicago Press.
Feyerabend, P. (1999a). *The conquest of abundance.* Chicago, IL: University of Chicago Press.
Feyerabend, P. (1999b). *Philosophical papers: Vol. 3. Knowledge, science and relativism.* Cambridge, England: Cambridge University Press.
Feyerabend, P. (2010). *Against method* (4th ed.; with an introduction by Ian Hacking). London, England: Verso.
Feyerabend, P. (2011). *The tyranny of science.* Cambridge, England: Polity Press.
Munévar, G. (Ed.). (1991). *Beyond reason: Essays on the philosophy of Paul Feyerabend* (Boston Studies in the Philosophy of Science). Dordrecht, Netherlands: Springer.
Oberheim, E. (2006). *Feyerabend's philosophy.* Berlin, Germany: Walter de Gruyter.
Preston, J. (1997). *Feyerabend: Philosophy, science and society.* Cambridge, England: Polity Press.
Preston, J., Munévar, G., & Lamb, D. (Eds.). (2000). *The worst enemy of science? Essays in memory of Paul Feyerabend.* Oxford, England: Oxford University Press.

Folk Psychology

Folk psychology (FP) is the commonsense understanding of other minds; it is the kind of psychology one does without any formal training. This entry reviews FP and the debates in it as well as recent developments and alternative theories on offer.

FP and related philosophical discussions have important repercussions on how to explain the way people routinely understand each other's beliefs and actions by employing pretheoretical, everyday psychological notions. In that sense, it is a social practice. So FP accounts are obviously important for the future of various social sciences.

Folk Psychological Notions

When we speak about what people "desire," "believe," "hope," or even "remember" and "see," we are appealing to *folk* (everyday pretheoretical) notions that, for the most part, were not introduced by scientists as either explanatory or predictive posits. The folk use these terms to make sense of behavior by putting it in context or identifying purported causes of behavior. In this way, we, the folk, attribute mental states to other people.

On the traditional narrow understanding of FP, the notions are limited to those attitudes that take a proposition as an object—that is, *propositional attitudes.* These are attitudes like *believe that* and *desire that,* such as "Poppy believes that the flowers are red." Attitudes that do not take propositions as their objects or instances of an attitude that don't have a proposition as an object, as in "Poppy desires cake," would not count as folk psychological attributions on this narrow view. A wide understanding of FP includes such attributions, as well as additional elements such as emotions and personality traits. Attributions such as "Poppy feels guilty" or "Poppy is considerate" fall within the scope of FP on this view.

Some attributions can be interpreted both widely and narrowly. Consider the term *sees,* which is sometimes translatable into *believes.* For example, "Poppy sees that there is a cake on the table" means "Poppy believes that there is a cake on the table." When the term *sees* takes an item as its object, it would not count as FP understood narrowly, as in "Poppy sees cake." The psychologists Michael Tomasello and Josep Call claim that chimpanzees understand seeing and can attribute perceptual states to others but that there is no evidence that they attribute beliefs, suggesting that chimpanzees may be folk psychologists on a wide, but not narrow, reading of FP. Another example is *knowing.* The same team, along with the psychologist Juliane Kaminski, does not hesitate to claim that chimpanzees know what others know, despite the fact that philosophers often define knowledge in terms of belief.

The FP-narrow concept is closely related to *theory of mind* (ToM), or the capacity to ascribe

propositional attitudes to others (also called *mind reading*). The psychologists David Premack and Guy Woodruff introduced the term in 1978 when they asked if the chimpanzee has a ToM, and the topic was taken up by the developmental psychologists Josef Perner and Heinz Wimmer. This research led to the conclusion that by age 4 human children are able to attribute propositional attitudes to others in terms of false beliefs. More recently, the psychologist Renée Baillargeon introduced nonverbal ToM tests, and some interpret the data as evidence that infants have a ToM before their first birthday. Despite the lack of experimental evidence for ToM in other animals, anecdotal evidence, the behavior of enculturated chimpanzees, and concerns about prior methodology leave open the possibility that other animals may share this capacity with humans.

The use of FP in science has been challenged in terms of its usefulness for individuating behaviors in psychology and neuroscience. For example, Angeline Lillard argues that FP-narrow concepts are not culturally universal, and Paul Churchland claims that FP-narrow concepts will be eliminated once we gain new and more precise notions from a mature neuroscience.

Theories of Folk Psychology

While the focus is on the FP practices of predicting and explaining behavior, the practices of FP are much broader than this, and they include everything from judging and coordinating behavior to teasing and flirting. Attributing FP notions alone is not sufficient for FP practices; what is needed is some understanding of how these notions relate with one another and with behavior.

The mind-reading debate between *Theory Theory* and *Simulation Theory* versions of FP was a matter of much discussion, though it has largely ended with many of the players accepting some hybrid view. These views are fully presented as their own entries in this encyclopedia.

Others working in FP eschew that debate—and the mind-reading label. Embodied, normative, and pluralistic views of FP stress the low cognitive demands for many FP practices. For example, Kristin Andrews suggests that the quotidian acts of prediction (of others' behavior or beliefs) involve a host of different cognitive processes, including inductive reasoning, trait attribution, and stereotype activation, and that these are distinct from belief attribution. Daniel Hutto suggests that such quotidian practices are examples of mind minding, or attending to the objects of intentional attention. Tadeusz Zawidzki suggests that mind shaping, or setting up regulative ideals that shape behavior, is behind our successful predictions.

Theories abound about the evolution of FP-narrow abilities in humans and the distribution of them among species. In his version of the Social Intelligence Hypothesis, Nicholas Humphrey argues that humans and perhaps other primates developed the ability to think about others' minds as a strategy for succeeding in a Machiavellian world of cheaters and thieves. Daniel Hutto's Narrative Practice Hypothesis suggests that it is having language, which allows us to develop and hear stories about people's reasons for their actions, that led to the uniquely human ability to understand others' actions in terms of their beliefs. Kristin Andrews argues that our understanding of others in terms of beliefs and desires rests on an understanding of social norms and a desire to explain norm violations and that these may be capacities we share with other species.

Kristin Andrews

See also Developmental Psychology; Embodied Cognition; Joint Attention and Social Cognition; Metacognition and Agency; Primatology and Social Science Applications; Simulation Theory; Theory Theory

Further Readings

Andrews, K. (2012). *Do apes read minds? Toward a new folk psychology.* Cambridge: MIT Press.

Baillargeon, R, Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Science, 14*(3), 110–118.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Science, 12*(5), 187–192.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78*(2), 67–90.

Hutto, D. D. (2008). *Folk psychological narratives: The sociocultural basis of understanding reasons.* Cambridge: MIT Press.

Hutto, D. D., & Ratcliffe, M. (Eds.). (2007). *Folk psychology re-assessed.* Dordrecht, Netherlands: Springer.

Lillard, A. S. (1998). Ethnopsychologies: Cultural variations in theories of mind. *Psychological Bulletin, 123*(1), 3–32.

Perner, J. (in press). Theory of mind: An unintelligent design. In A. M. Leslie & T. C. German (Eds.), *Handbook of "theory of mind."* Piscataway, NJ: Lawrence Erlbaum.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral & Brain Sciences, 1*(4), 515–526.

Zawidzki, T. W. (in press). *Mindshaping*. Cambridge: MIT Press.

Formal Epistemology

Formal epistemology is the application of formal methods (logic, probability theory, and algebra, etc.) to an analysis of the operational properties of knowledge and related epistemic phenomena such as belief and information. Just as in the physical sciences we use formal methods to construct mathematical models of physical phenomena, in formal epistemology, we use formal methods to construct mathematical models of epistemic phenomena.

The earliest example of *epistemic logic* is to be found in Jaakko Hintikka's 1962 work *Knowledge and Belief,* itself influenced by Georg Henrik von Wright's 1951 *Essay in Modal Logic*. The starting point for epistemic logic is to encode knowledge and belief with the *modal operators K* and *B*. In this case, *agent α knows that A* and *agent α believes that A* will be encoded syntactically by (1) and (2), respectively:

1. $K_\alpha A$
2. $B_\alpha A$

If something is known, then it is true (on account of knowledge being factive). It is also believed (on account of knowledge implying belief), hence its negation is not believed. Thus, we have the axioms (3), (4), and (5):

3. $K_\alpha A \rightarrow A$.
4. $K_\alpha A \rightarrow B_\alpha A$.
5. $K_\alpha A \rightarrow \neg B_\alpha \neg A$.

Furthermore, axioms may be added in order to account for additional epistemic capacities. Note that the converse of (5), $\neg B_\alpha \neg A \rightarrow K_\alpha A$, does *not* hold, at least not without the additional assumption that agent α is maximally opinionated. Hence, *K* and *B* are not *duals* in the sense of other modal operators such as *necessity* $\Box$ and *possibility* $\Diamond$, where we have it both that $\Box A \leftrightarrow \neg \Diamond \neg A$ and $\Diamond A \leftrightarrow \neg \Box \neg A$. However, *K* and *B* do share some important properties with $\Box$ and $\Diamond$, respectively, namely, that the first member of each pair is a universal modality and the second member of each pair is an existential modality. To understand universal and existential modalities, we need to move beyond the syntax of epistemic logic to its semantics.

Modal operators may have their semantics specified by *possible worlds,* as first discussed by Saul Kripke. Possible worlds are simply abstractions of different ways the universe might have been. In basic epistemic logic, a *model* **M** is a triple consisting of a set of possible worlds *W,* a binary accessibility relation $\rightarrow_\alpha$ between members of *W,* and a valuation relation $\Vdash$ that assigns truth-values to propositions at worlds. In this case, *K* and *B* will have the following evaluation conditions with respect to **M** (where $w_@$ is the actual world).

6. $w_@ \Vdash K_\alpha A$ *iff* for all $w_i \in \mathbf{M}$ such that $w_@ \rightarrow_\alpha w_i$, $w_i \Vdash A$.
7. $w_@ \Vdash B_\alpha A$ *iff* for some $w_i \in \mathbf{M}$ such that $w_@ \rightarrow_\alpha w_i$, $w_i \Vdash A$.

It is in the sense specified by (6) and (7) that *K* and *B* are respectively universal and existential modalities, since *K* requires that *A* is true at *all* worlds accessible by α, while *B* requires that *A* is true at *some* worlds accessible by α. This is as we would expect, since belief, unlike knowledge, allows for the possibility that the proposition falling within its scope be false.

Contemporary research into epistemic logics and related areas is vast. One direction of expansion involves the switch from possible worlds to *information states*. Unlike possible worlds (which are complete and consistent), information states increase the modeling flexibility of the resulting logical systems by representing incomplete and/or inconsistent information. This allows for a very fine-grained analysis of real-world epistemic scenarios. Another direction of grain increasing is the distinction between *implicit* and *explicit* beliefs.

Perhaps the most important recent development in the logical area of formal epistemology is the move from *static* to *dynamic* epistemic logics. Dynamic

epistemic logics allow us to model the epistemic actions that facilitate the flow of information that underpins knowledge and belief. Information may flow between an agent and her brute physical environment, in which case the relevant epistemic actions will be *observations;* between two or more agents, in which case the relevant epistemic actions will be *announcements;* between different states of explicit awareness of a single agent, in which case the relevant epistemic actions will be *inferences;* and more.

Both static and dynamic epistemic logics may have their semantics given by information states and the relations between them. But a collection of objects with relations between them will correspond to an *algebra*. As such, there is a flourishing research program into algebraic models of epistemic phenomena, which extends into models carried out in extremely powerful mathematical frameworks such as *quantum logics* and *category theory*.

All the formal epistemic frameworks introduced so far have been *qualitative*. They allow for comparisons and combinations of bodies of stronger or weaker beliefs, more or lesser knowledge, more or less information, and so forth. In short, we may *order* such epistemic phenomena. In order to measure it—to impose discrete numerical values or weights to beliefs or bodies of information—we move from the qualitative frameworks of formal epistemology to the *quantitative* ones.

The canonical quantitative framework in formal epistemology is based upon *conditional probabilities*. The *conditional probability function* is $P(B|A)$, where $P(A) > 0$, given in (8):

8. $P(B|A) = P(A \cap B)/P(A)$.

Bayesian epistemology is a family of approaches (Fitelson, in press) founded upon *Bayes's theorem* (9), which, following immediately from (8), is a fundamental theorem of conditional probability. Bayes's theorem calculates the conditional probability (P) of a hypothesis (H) given some evidence (E), written as $P(H|E)$, as follows:

9. $P(H|E) = P(H)P(E|H)/P(E)$.

The probability measure is then taken as a *normative prescription* on your degree of belief. The rational degree of belief in H given one's total evidence E is just $P(H|E)$, calculated as stipulated by (9). Following Frank Ramsay, one's degree of belief is taken to be the odds that one would take on a bet that the proposition(s) encoded by the hypothesis is true, and one's total evidence is whatever evidence it is that caused one's belief to have the degree of confidence that it enjoys.

An important consequence of this shown by Ramsay is the Dutch Book theorem: If one is prepared to adjust one's betting pattern so that it maps onto one's degrees of belief, then these degrees of belief are governed by the probability calculus if and only if one cannot be outwitted by a sequence of bets constructed by any betting opponent, no matter how cunning.

Quantitative belief revision is underpinned by *conditionalization*. Conditionalization states that one's degrees of belief should map onto the probability calculus and that the revision of one's beliefs should proceed according to (8) or (9). In other words, in light of our observing B, our degree of belief in A specified by $P(A)$ should be adjusted according to the basic tenets of the probability calculus. Hence, belief revision is another *dynamic* turn in formal epistemology.

An important theorem for quantitative belief revision is *de Finetti's theorem*. Informally, the theorem's contribution is that granting certain conditions, as two agents increase their evidence bases with respect to an event series, their degrees of belief with respect to the likelihood of a future event in the series will meet.

Like research into the qualitative logical aspects of formal epistemology, research into the quantitative decision-theoretic frameworks of formal epistemology is extremely active. Normative quantitative frameworks for rational decision making group under Decision Theory. A powerful research subarea is Social Choice Theory, which contributes to sociology by providing analyses of group decision-making scenarios, such as voting outcomes.

Qualitative and quantitative approaches to formal epistemology merge and intersect at many places, both with models of belief aggregation and models of the information flow that underpins such beliefs. Game theory is a formal epistemic framework that provides both qualitative and quantitative models of a range of epistemic phenomena.

Formal epistemology interacts with a wide range of research areas. Such areas include the philosophy of science, with models of theory change and other belief revision frameworks such as AGM Theory (after the initials of the authors who established the

field—Alchourron, Gardenfors, and Makinson); artificial intelligence, with models of inference; theoretical computer science, with models of reasoning resources via linear logics; and more. Given that formal epistemology models epistemic phenomena and given the ubiquitous presence of such phenomena across research into rational behavior, this is as we should expect it to be.

Sebastian Sequoiah-Grayson

See also Artificial Intelligence; Bayesianism, Recent Uses of; Decision Theory; Epistemology; Game-Theoretic Modeling; Induction and Confirmation; Social Choice Theory

Further Readings

Abramsky, S. (2008). Information, processes, and games. In P. Adriaans & J. F. A. K. van Benthem (Eds.), *Handbook on the philosophy of information* (pp. 483–550; Handbook of the Philosophy of Science Series). Amsterdam, Netherlands: North-Holland.

Alchourron, C. E., Gardenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic, 50,* 510–530.

Baltag, A., & Smets, S. (2008). Probabilistic dynamic belief revision. *Synthese, 165*(2), 179–202.

de Finetti, B. (1980). Foresight: Its logical laws, its subjective sources. In H. E. Kyburg Jr. & H. Smokler (Eds.), *Studies in subjective probability* (pp. 53–118). Huntington, NY: Krieger.

Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (2004). *Reasoning about knowledge.* Cambridge: MIT Press.

Fitelson, B. (in press). Contrastive Bayesianism. In M. Blaauw (Ed.), *Contrastivism in philosophy.* London, England: Routledge.

Hartmann, S., Martini, C., & Sprenger, J. (Eds.). (2010). *Formal modelling in social epistemology* [Special issue]. *Logic Journal of the IGPL, 18*(2).

Kaplan, M. (1998). *Decision theory as philosophy.* Cambridge, England: Cambridge University Press.

Kripke, S. (1963). Semantical analysis of modal logic. *Zeitschrift fur Mathematichs Logik und Grundlagen der Mathematik, 9,* 67–96.

Levi, I. (2004). *Mild contraction: Evaluating loss of information due to loss of belief.* Oxford, England: Clarendon Press.

Liu, F. (2011). *Reasoning about preference dynamics.* Dordrecht, Netherlands: Springer.

Ramsay, F. (1931). *The foundations of mathematics and other essays.* New York, NY: Humanities Press.

Sadrzadeh, M., & Dyckhoff, R. (2010). Positive logic with adjoint modalities: Proof theory, dynamics, and reasoning about information. *Review of Symbolic Logic, 3*(3), 351–373.

Seligman, J. (2009). Channels: From logic to probability. In G. Sommaruga (Ed.), *Formal theories of information: From Shannon to semantic information theory and general concepts of information* (pp. 193–234; LNCS 5363). New York, NY: Springer.

Sequoiah-Grayson, S. (2010). Epistemic closure and commutative nonassociative residuated structures. *Synthese.* Advance online publication. doi:10.1007/s11229–010–9834-z

van Benthem, J. F. A. K. (2011). *Logical dynamics of information and interaction.* Cambridge, England: Cambridge University Press.

van Ditmarsh, H., van der Hoek, W., & Kooi, B. (2008). *Dynamic epistemic logic* (Synthese Library, Vol. 337). Dordrecht, Netherlands: Springer.

Velazquez-Quesada, F. R. (2011). *Small steps in dynamics of information* (ILLC Dissertation Series, DS-2011–02). Amsterdam, Netherlands: Institute for Logic, Language and Computation.

von Wright, G. H. (1951): *An essay in modal logic.* Amsterdam, Netherlands: North-Holland.

Weirich, P. (2004). *Realistic decision theory: Rules for non-ideal agents in non-ideal circumstances.* New York, NY: Oxford University Press.

Foucault's Thought

Michel Foucault was a French philosopher and historian of ideas who was born in Poitiers in France in 1926 and died in Paris in 1984. Since his death, his work has had a steadily increasing impact across the social sciences and humanities, generating new research methodologies, new areas of empirical interest, and a whole panoply of theoretical concepts. Foucault produced some 11 books during his lifetime, and a collection of 364 of his shorter writings was published in 1994. From 1970 to 1984, in his capacity as Professor of Systems of Thought at the research institution the Collège de France, he also produced an annual series of lectures reporting on his research. These lectures have gradually

appeared in print since 1997. This entry provides an overview of Foucault's overall philosophy and methodology, looks at key concepts in his work, and provides descriptions of his best-known books.

Philosophy and Methodology

Central to Foucault's work is the idea that all human experience, and knowledge and organization of the social and physical world are historical, from the most esoteric of philosophical abstractions to the most mundane of everyday activities. Notions that had hitherto more commonly been regarded as unchanging givens are historicized in Foucault's work—objects that include power, madness, sexuality, delinquency, the normal and the pathological, science, truth, knowledge, the subject (self), and ethics. For Foucault, not only is human existence entirely historical but also all social, cultural, and material practices embody processes of human thought. Knowledge and our perceptions of truth are not static but the direct result of our ever-changing interaction with our physical environment and with each other. Such an assumption opens the way for a variety of explanatory approaches that use detailed empirical work as a support in areas previously regarded as the province of speculative philosophy.

A strong social justice agenda also lies at the heart of Foucault's historico-philosophical approach. Foucault was part of a more general post–World War II social and intellectual movement that drew attention to the experiences of marginalized and colonized populations and the social strictures to which they and also more mainstream members of the social body were subject. Foucault also sought to examine bodies of knowledge that had been neglected and silenced by triumphalist accounts of scientific, technical, and social "progress."

If the subject matter of Foucault's writing often deals with the darker side of human existence (madness, illness, crime, abnormalities of all kinds, and the multiple restrictions around sexual practice), he is a fundamentally optimistic thinker in that he believes that by exposing this dark side, people will find certain aspects of the accepted status quo intolerable and be encouraged to develop more socially just practices. For all this, Foucault's view is that we will probably never reach that utopian point, promised by so many 19th-century post-Enlightenment thinkers, where all our social ills have been eradicated. Each solution brings an attendant set of new problems, which, in turn, need sustained work and reflection.

Foucault maintained a fairly consistent methodological approach throughout his career, even if he changed his descriptive labels. He deliberately sets out to oppose certain traditional notions and assumptions that were common currency in 19th- and 20th-century history and philosophy. These assumptions include belief in ideas of historical continuity, progress, and a universal unchanging subject of consciousness and action. He also challenges the notion that unchanging objects are "discovered" by this universal subject. Instead, what we see is a constant process of historical redefinition of both subjects and objects and a complex set of negotiations and relations between these terms that Foucault described in terms of knowledge, power, and truth.

Knowledge, Power, and the Subject

If Foucault's approach is held together by an overall philosophical and methodological unity, there are also three distinct periods in his work, marked by differences in emphasis, subject matter, and terminology. The first period, which is from 1954 to 1970, is characterized by a focus on the history and philosophy of science, in particular the medical sciences and the human sciences. To distinguish his work from standard approaches in the history of ideas, Foucault uses the word *archaeology* to describe his methodology during this period.

During the second phase of his work, from 1970 to 1981, he changed both the subject matter of his analyses and the emphasis of his methodology. While still working with the same historical period (from the mid 18th century to the mid 19th century, with some brief excursions into earlier European history), Foucault focused on the history of techniques designed to organize and govern populations. These techniques were deployed in institutions such as prisons, schools, hospitals, asylums, armies, and the state. He adopted a new term, *genealogy* (borrowed from Nietzsche), to describe his method. If archaeology focused on discourses and the ordering of knowledge systems, genealogy shifted the focus to a stronger and more explicitly detailed emphasis on the role of relations of power in the organization

of knowledge and institutions. In the third phase, during the 1980s, Foucault surprised his readers by shifting his focus away from broad sociocultural, institutional, and political arrangements to a focus on historical processes relating to the way individual subjects, the self, and ethical systems were constituted in classical antiquity and the early Christian era.

Foucault's Major Works

1954–1969

Foucault's first publications in the 1950s dealt with the history of psychology and psychiatry, and his first major work, which appeared in 1961, *History of Madness,* essentially rewrote standard histories of psychiatry. In this book, Foucault addresses the changing medical, economic, political, institutional, philosophical, ethical, artistic, and literary practices, which, from the 13th century onward, in Western culture, helped define madness as a cultural, social, and scientific category. He challenges traditional accounts demonstrating that the birth of modern psychiatry in the 19th century was not the result of the inevitable progress and victory of enlightened science over ignorance and superstition. It was instead the result of changes in what was accepted as valid knowledge in relation to a complex biological and social phenomenon as well as changes in practices relating to its institutional and political management.

In *The Birth of the Clinic* (1963), Foucault once again takes issue with traditional accounts in his account of the birth of clinical medicine between 1769 and 1825 in France. Standard histories held that a dubious reliance on superstition and ancient texts was replaced by a more enlightened scientific focus on visible and empirically observable symptoms. For Foucault, however, it is more a matter of a shift in the rules of what counted as valid and true knowledge. New notions of illness (the object) and what constituted medical practice as embodied in the doctor (the "gazing" or observing subject) were developed along with new institutions that recodified the identities and relationships among doctors, patients, and illnesses. The notions of the "gaze" and visibility have found considerable fortune in the secondary literature and tie in with Foucault's later work in *Discipline and Punish* about the growth of a disciplinary society that uses surveillance as one of its primary mechanisms of social control.

Perhaps Foucault's most obvious and direct engagement with the human and social sciences can be found in his 1966 work *The Order of Things,* which examines the origins of the modern disciplines of economics, linguistics, and biology. The book also includes a final chapter that deals with the formation of the disciplines of sociology, ethnology, history, and psychoanalysis.

If specialists remain interested in the intricacies of the arguments in *The Order of Things,* what has generally been popularly retained from this book are two notions. The first of these is the idea that the humanist subject, far from being universal, is a 19th-century creation. The second notion is the "episteme," similar in some respects to Thomas S. Kuhn's notion of the paradigm. It describes a configuration of scientific knowledge lasting roughly around 150 years that sets the rules for the production of knowledge regarded as valid.

Foucault's next work, *The Archaeology of Knowledge,* published in 1969, addresses a notion that has become familiarly associated with his work—namely, "discourse." At the most general level, Foucault uses this term to designate the material traces left by words in history: Discourse and words are not just transparent signposts to an external reality, they have their own substance and rules. Discourse is also a structured way of organizing particular forms of knowledge (e.g., psychiatry, economics, and biology) and a way of speaking about the world.

1970–1984

Like many other French intellectuals, Foucault was politicized by the social unrest and student and worker protests of 1968, and his inaugural speech at the Collège de France introduced specific references to the notion of power. From this point onward, Foucault's interest in the relation between knowledge and power took a practical as well as a theoretical turn. Demonstrating that he was no armchair philosopher, he took part in committees and demonstrations, as well as signing numerous petitions to support the rights of prisoners, health care workers, and immigrants and, later, Polish trade unionists and boat people. During the 1970s, Foucault introduced a number of productive concepts centering on power, which have been widely adopted in the secondary literature. These include

power-knowledge, discipline and surveillance, biopower, and governmentality.

The year 1975 saw the publication of *Discipline and Punish,* a history of the birth of the prison as the modern punishment of choice for criminal behavior and an account of what Foucault describes as the birth of the disciplinary society. Focusing on the period between 1757 and 1838, he argues that there was a move away from corporal punishment that took place in the public arena to punishment that focused on the deprivation of liberty and rehabilitation of socially deviant identities. Foucault proposes that the 18th-century British philosopher Jeremy Bentham's model prison, the Panopticon, where prisoners were watched round the clock by unseen guards or members of the public, serves as a metaphor for modern society, which relies on mechanisms of surveillance and discipline operationalized by a variety of institutions, including schools, prisons, hospitals, and armies.

In *The History of Sexuality: Vol. 1. The Will to Know* (1976), Foucault reverses the commonplace that European sexuality underwent a massive repression between the end of the 17th and the end of the 19th century, arguing that, far from repression, there was an immense expansion of knowledge around sexuality. This knowledge was one element in the control of the health and life cycle of populations so as to maximize the power and cohesion of modern industrialized nation-states. Foucault uses the term *biopower* to characterize this arrangement.

Foucault also developed his theories of power across a number of other shorter publications during the 1970s. His ideas challenge mainstream theories such as the idea that power is principally a property of the state and the ruling classes and, further, that power is by definition repressive. Instead, Foucault argues that power only exists as a relationship between different parties. If Foucault started with a fairly broad notion that opposed relations of power and resistance, eventually he adopted the position that all those involved in power relations necessarily had some degree of freedom of choice. When that freedom is shut down by violence or slavery, it is no longer a relationship of power. In short, power has its limits as a description and explanation of social relations.

In his 1977–1978 lectures, Foucault introduced the notion of governmentality, a notion that has become widely applied. If initially he used the term in relation to the rationalization of the exercise of state power and administrative management of populations, he gradually broadened it in the early 1980s to encompass an analysis of how individual subjects are governed and govern themselves and the techniques that allowed people to work on themselves and in relation to others.

Foucault's last two books, Volumes 2 and 3 of *The History of Sexuality: The Use of Pleasure* and *The Care of the Self,* published in 1984 a month before his death, examine techniques of self-formation in the work of the Ancient Greeks and Romans and early Christians. Of particular note is his development of a schema he uses to describe how people have historically constructed themselves as moral and ethical subjects. He systematizes this process into four levels: (1) the part of the self to be worked on (e.g., desires, acts, feelings), (2) the authority that makes an individual recognize his or her moral obligations (the Bible, rationality, etc.), (3) the work people do to transform themselves (meditation, diet, physical training, etc.), and (4) the desired outcome (everlasting life, salvation, self-mastery, etc.).

The sheer breadth of Foucault's interests and the flexibility of the philosophical and methodological tools he offers continue to make his work a fertile starting point for researchers across a wide diversity of theoretical and applied disciplines.

Clare O'Farrell

See also Contemporary French Philosophy and the Social Sciences; Discourse Analysis; Genealogy; Governmentality and Regime; Postmodernism; Power; Sexuality

Further Readings

Foucault, M. (1970). *The order of things: An archaeology of the human sciences* (A. M. Sheridan-Smith, Trans.). London, England: Tavistock. (Original work published 1966)

Foucault, M. (1972). *The archaeology of knowledge* (A. M. Sheridan-Smith, Trans.). London, England: Tavistock. (Original work published 1969)

Foucault, M. (1973). *The birth of the clinic: An archaeology of medical perception* (A. M. Sheridan-Smith, Trans.). London, England: Tavistock. (Original work published 1963)

Foucault, M. (1990). *The history of sexuality: Vol. 3. The care of the self* (R. Hurley, Trans.). Middlesex, England: Penguin. (Original work published 1984)

Foucault, M. (1991). *Discipline and punish: The birth of the prison* (A. M. Sheridan-Smith, Trans.). Middlesex, England: Penguin. (Original work published 1975)
Foucault, M. (1992). *The history of sexuality: Vol. 2. The use of pleasure* (R. Hurley, Trans.). Middlesex, England: Penguin. (Original work published 1984)
Foucault, M. (1998). *The history of sexuality: Vol 1. The will to knowledge* (R. Hurley, Trans.). Middlesex, England: Penguin. (Original work published 1976)
Foucault, M. (2006). *History of madness* (J. Murphy & J. Khalfa, Trans.). London, England: Routledge. (Original work published 1961)

Frankfurt School and Critical Social Theory

Critical theory comprises a loosely associated group of philosophers, sociologists, political scientists, artists, and theologians who write incisive critiques of Western philosophy, ethics, law, and culture. These thinkers belong to the heritage inspired by the Institute of Social Research, founded in the early 1920s at the University of Frankfurt by Carl Grünberg and Felix Weil. The Institute had emerged as an academic and intellectual response to the collapse of the German political, social, and economic orders in the period immediately following World War I. Inspired by the social theory of the so-called Western Marxists, such as Georg Lukàcs (1885–1971), Karl Korsch (1886–1961), and Antonio Gramsci (1891–1937), the members of the early Institute examined certain Marxist tenets that they thought could be taken up and modified to confront these problems.

By the 1930s, Max Horkheimer (1895–1973) became the Institute's leader and chief theoretician. In his influential essay of 1937, "Traditional and Critical Theory," he outlined a broad program of theoretical and empirical-practical research for the Institute. But since most of its members were Jewish, the Institute was dissolved during the prelude to World War II, forcing most of the members and its associates into exile. Max Horkheimer, Walter Benjamin (1892–1940), Theodor Adorno (1903–1969), Erich Fromm (1900–1980), Herbert Marcuse (1898–1979), and Leo Löwenthal (1900–1993) fled Germany, with all but one, Benjamin, ending up in the United States. After the war, the Institute was formally reassembled in Frankfurt and continues to this day. Jürgen Habermas (1929–) and Karl Otto-Apel (1922–) emerged in the 1960s as the most representative and influential members of what is considered the second generation of critical theory. Now, a third generation, representing scholars from around the world, is emerging and continuing in the same critical direction.

Theoretical Principles

Critical theorists were united initially in opposition to the prevailing positivism of the day. On their account, positivism held that social, economic, and political questions could be resolved solely by objective, and thus uncritical, means. For example, historical studies were dominated by the conviction that a historian could in principle find invariant laws governing historical events. The Institute's work, though informed by science, stipulated that scientists must be critically aware of the historical conditions of their own specific research. Critical theorists were inspired in great part by Max Weber's (1864–1920) *Verstehende* sociology, according to which the researcher's own interpretive perspective is considered constitutive of what is studied. Critical theorists also adopted elements of Freudian psychology, both its distinction between the pleasure (*eros*) and death (*thanatos*) drives and its utilization of the structural models of the id, ego, and superego. These conceptual schemes became a template on the basis of which sociological phenomena could be analyzed. With them, critical theorists analyzed dominant social ideologies latent in the family, the mass media, and other cultural and political systems.

Critical theorists also found themselves in opposition to phenomenology and existentialism, two philosophical movements also gaining influence in the interwar period. The phenomenological method, developed principally by Edmund Husserl (1859–1938), was intended to be a rigorous science of consciousness that reduced everyday experience to a series of noematic essences. Critical theorists saw in the method, however, merely a way to keep dominant class interests protected by leaving all power of interpretation to social and economic elites. They saw in existentialism—a method of reduction not to essences but to existence itself—the same adoption of a structure of meaning and truth immune from comprehensive historical social critique. In response, Benjamin developed, in his *Origins of*

German Tragic Drama (1928), a theory of interpretation derived from the notion of ideas. Ideas emerge not from a conceptual analysis of things but by a formation of phenomena into "constellations." The relation between ideas and phenomena is no longer understood as a mutual participation grasped by an intuition. Rather, ideas are articulated through "allegorical experience" that expresses a form while grasping its origin simultaneously within a historical epoch. In his *The Work of Art in the Age of Mechanical Reproduction* (1935), Benjamin indicated that, when understood in this way, the expressions of artworks, particularly film, could serve powerful political ends.

The principal method employed by the Frankfurt school was, and continues to be, that of an *immanent critique* of social phenomena. Immanent critique gives up on a correspondence theory of truth, according to which the mind is understood to conform itself to a fixed reality, and instead employs only targeted inquiries into a historical lineage preceding a current mindset or social practice. For example, critical theorists argue that the historical development of capitalism parallels the growth of a dominating and inescapable system of instrumental reason. Early critical theorists like Benjamin, as well as Horkheimer and Adorno in their seminal 1943 work *The Dialectics of Enlightenment,* held that the contemporary world teetered on the edge of catastrophe and escape from it could come only from a complete withdrawal from instrumental (enlightenment) rationality. Horkheimer and Adorno employed a critique of the "culture industry" that exposed the way in which literature, film, music, and the fine arts had all been co-opted by the profit motive of capitalist systems. Though Horkheimer and Adorno remained pessimistic regarding the possibility of any lessening of social domination, Marcuse in his 1964 book *One-Dimensional Man* argued that science and technology could be oriented away from capitalism's "false needs" toward nonrepressive and liberatory ends.

When the Institute was revitalized in Germany after World War II, critical theorists confronted social conditions no longer of economic crisis and war but rather of robust economic recovery and relative political stability. Now their immanent critique concentrated rather on the problems of social conformity and psychological repression characteristic of this prosperity. In this period, Benjamin's work on aesthetics and ethics in the 1920s and 1930s became more salient. Inspired by this, Adorno's publication of his aphoristic ethical work *Minima Moralia* (1956) contained a broad critique of everyday social mores.

In the 1960s, Adorno turned from his cultural and ethical critique to a more theoretical direction. His *Negative Dialectics,* published in 1966, traces the way in which both the critical thinking of Kant and the positive dialectics of Hegel set the stage for a philosophical thinking ordered toward the ideal of freedom. However, neither project succeeded, since both failed to challenge the social ideal, operative in their times, of a "free" bourgeois subject acting on the basis of its own self-interest. Instead, Adorno defended a nonidentity thinking that, while emergent from the prior development of identity thinking and its emphasis on subjectivity, shifts toward the "primacy of the object." Nonidentity thinking considers the enlightenment impulse toward universalization as inherently incomplete. Adorno envisioned not a positive structural metaphysics where form dominates matter but a "rescuing metaphysics" in which particulars, while in dialectical tension with universals, resist subsumption by them.

By the late 1960s, critical theorists had earned a measure of public notoriety. In the United States, Marcuse's works were read by students and public intellectuals supporting the idea of a counterculture and social revolution. In Germany, Adorno and Habermas were solicited by students for support for numerous leftist causes. However, both Adorno and Habermas refused to support the tactical proposals of the students, responding instead only with theoretic analyses that they thought would keep the movements from becoming reactionary or sensationalistic. This led to bitter disappointments on the part of many of the students. Adorno's untimely death in 1969 and Habermas's transfer from the University of Frankfurt to the Max Planck Institute in Starnberg in 1971 ended the strained dialogue.

Meanwhile, a new theoretical direction was emerging in Habermas's work. He characterized the critique of the Enlightenment by his former mentor Adorno as a mere "philosophy of consciousness" that, as such, was bereft of emancipatory potential. In his 1981 work *The Theory of Communicative Action,* Habermas broadened the scope of the theory of rationality from the analysis of subjectively produced instrumental actions to a comprehensive

mapping of all intersubjective communication. Initially, he defended a consensus theory of truth, on the basis of which truth was understood to be identical with its mode of intersubjective verification by the assent of all those potentially affected by the norm in question. But later, in his *Truth and Justification* (1999), he turned from this solely verificationist model toward a more realist model that distinguishes between verification processes and invariant truth.

Drawing from Karl-Otto Apel's work on formal pragmatics—based on the grasp of how the absence of performative contradictions gives evidence of a rational procedure—Habermas also developed a verificationist moral theory, termed *discourse ethics*. Discourse ethics presumes that just as we garner valid scientific conclusions from dialogue and discussion, so do we also draw the rightness of moral norms from discourse among those affected by them. His early version of discourse ethics appealed to an "ideal speech situation," from which perspective all affected by the eventual norm could be guided toward consensus. In later versions, however, he spoke of how discourse is guided rather by "idealizations," such as that of an orientation toward universalization and inclusiveness on the part of all those affected in any way by the norm in question. Regardless, he never urged that any absolute rightness of a norm could be derived from discourse. Moreover, he relied on the legal theory of Klaus Günther to develop criteria by which verified moral norms could be applied in unique situations.

One critic of discourse ethics was Albrecht Wellmer, who questioned its limiting of truth and rightness only to the redemption of validity claims in a formal pragmatic structure. Other critics similarly rejected its exclusion of contextual conditions—such as gender, social status, and psychological considerations. But Habermas developed, in *Between Facts and Norms* (1992), a political and legal theory on the premise that laws are necessitated by the inability of moral and ethical discourse alone to provide norms that cover all situations and contingencies. He argues that the discourse principle itself can reveal how individual rights (private autonomy) and institutionalized but changeable public laws (public autonomy) mutually presuppose one another so as to then guarantee the efficacy of principles of justice for all citizens.

The Third Generation of Critical Theory

Third-generation critical theorists, from a number of diverse academic disciplines, are today moving critical-theoretic analysis in many new directions.

Responding to Foucault's analysis, Axel Honneth traces social pathologies to struggles for power and recognition among individuals and groups. He shifts from structural to individual and intersubjective bases of pathologies. Thomas McCarthy, whose translations of Habermas's work facilitated the spread of critical theory in North America, has turned critical theory toward an analysis of the problems of race. David Ingram has done similar work. Rainer Forst analyzes a critical-theoretic model of justice able to address the persistence of social and political intolerance in contemporary societies. Lambert Zuidervaart shows how Adorno's work contains a critical social theory able to deal with issues of economic domination. Nancy Fraser and Seyla Benhabib focus on the inability of prior critical theory to take sufficient account of women's and minorities' unique difficulties with regard to communicative and recognitional access. Joseph Heath has taken up rational choice theory in an effort to make the rationality critique of critical theory more salient. Cristina Lafont works in the area of moral and political theory, particularly regarding the issues of political illegitimacy. Similar work has been done by David Rasmussen regarding violations of human rights. James Bohman works on social analysis that aims to take into account the new world situation brought about by increasing globalization. William Rehg has analyzed extensively both discourse and argumentation theory, particularly that which would inform the social sciences. In the area of psychoanalysis, Joel Whitebook argues that a theory of sublimation can provide an analysis of genesis and validity that can avoid the difficulties of reliance on idealizations. Finally, several theologians, such as Helmut Peukert, Thomas Schmidt, and Francis Schüssler Fiorenza, employ a critical-theoretic analysis of dogmatic forms of theology.

In the future, critical theory will undoubtedly continue to develop both theoretical models of emancipative thinking and practical ways of living aimed to challenge—as Horkheimer famously called them—*threats to freedom*.

James Swindal

See also Alienation: From Philosophy to Sociology; Communicative Action Theory; Dialectic, in the History of Philosophy; Dialectic, in the Social Sciences; Enlightenment, Critique of; Neo-Marxism; Transcendental Pragmatics; Weber's *Verstehende* Approach

Further Readings

Adorno, T. (1973a). *Negative dialectics*. New York, NY: Seabury Press.

Adorno, T. (1973b). *Philosophy of modern music*. London, England: Sheed & Ward.

Adorno, T. (1974). *Minima moralia: Reflections from damaged life*. London, England: New Left Books.

Adorno, T. (2002). *Metaphysics: Concept and problems*. Palo Alto, CA: Stanford University Press.

Adorno, T., & Horkheimer, M. (1972). *The dialectics of enlightenment*. New York, NY: Herder & Herder.

Apel, K.-O. (1984). *Understanding and explanation*. Cambridge: MIT Press.

Benhabib, S. (1986). *Critique, norm, and utopia*. New York, NY: Columbia University Press.

Benjamin, W. (1968). *Illuminations: Essays and reflections*. New York, NY: Schocken.

Benjamin, W. (1979). *Reflections*. New York, NY: Harcourt Brace Jovanovich.

Buck-Morss, S. (1977). *The origins of negative dialectics*. New York, NY: Free Press.

Fraser, N. (1989). *Unruly practices: Power, discourse, and gender in contemporary social theory*. Minneapolis: University of Minnesota Press.

Habermas, J. (1968). *Knowledge and human interests*. Boston, MA: Beacon Press.

Habermas, J. (1987). *Theory of communicative action* (2 vols.). Boston, MA: Beacon Press. (Original work published 1981)

Habermas, J. (1989). *Moral consciousness and communicative action*. Cambridge: MIT Press.

Habermas, J. (1992). *Between facts and norms*. Cambridge: MIT Press.

Honneth, A. (1993). *Critique of power*. Cambridge: MIT Press.

Horkheimer, M. (1972). Traditional and critical theory. In *Critical Theory: Selected Essays* (pp. 188–243). New York, NY: Herder & Herder.

Jay, M. (1973). *The dialectical imagination: A history of the Frankfurt School and the Institute of Social Research, 1923–1950*. Boston, MA: Little Brown.

Marcuse, H. (1955). *Eros and civilization: A philosophical inquiry into Freud*. Boston, MA: Beacon Press.

Marcuse, H. (1964). *One-dimensional man*. Boston, MA: Beacon Press.

McCarthy, T. (1991). *Ideal and illusions: On reconstruction and deconstruction in contemporary critical theory*. Cambridge: MIT Press.

Rasmussen, D., & Swindal, J. (Eds.). (2004). *Critical theory* (4 vols.). London, England: Sage.

Rehg, W. (2009). *Cogent science in context: The science wars, argumentation theory, and Habermas*. Cambridge: MIT Press.

Wellmer, A. (1991). *The persistence of modernity: Essays on aesthetics, ethics, and postmodernism*. Cambridge: MIT Press.

Wiggershaus, R. (1994). *The Frankfurt school: Its history, theories, and political significance*. Cambridge: MIT Press.

Zuidervaart, L. (2007). *Social philosophy after Adorno*. Cambridge, England: Cambridge University Press.

Free Will, Philosophical Conceptions of

This entry charts the philosophical terrain held by different positions on free will in relation to moral responsibility.

Free Will and Moral Responsibility

The problem of free will and moral responsibility arises from a tension between two powerful considerations. On the one hand, we human beings typically believe that we are morally responsible for our actions, which would require us to have some type of free will. On the other hand, scientific and theological views about the nature of the universe motivate causal determinism, and our having free will and determinism being true are in apparent conflict.

The history of philosophy records three standard reactions to this tension. *Compatibilists* maintain that it is possible for us to have the free will required for moral responsibility even if the universe is deterministic. Others argue that determinism and the contention that we have this sort of free will cannot both be true—they are *incompatibilists*—but they resist the reasons for determinism and affirm that we do have this free will of this kind. They advocate the *libertarian* position. *Hard determinists* are also

incompatibilists, but they accept that determinism is true and that we lack the free will required for moral responsibility. Especially since David Hume's discussion of these issues, the concern about the existence of the sort of free will required for moral responsibility has been extended to whether it is compatible with the *indeterminacy* of actions. This development has challenged the value of the threefold classification, despite its persistence in the contemporary debate. Some, like Galen Strawson, maintain that the free will required for moral responsibility is incompatible not only with determinism but with indeterminism as well.

Libertarianism

Libertarians prior to the modern period include the Epicurean author Lucretius, the early Christian theologian Origen, and among the scholastics, John Duns Scotus, William of Ockham, and Francisco Suarez. More recent times have witnessed the explicit differentiation of three major versions of libertarianism. In *event-causal libertarianism,* actions are caused solely by prior events such as an agent having a desire at a time, and indeterminacy in the causation of actions by events of this kind is held to be a critical requirement for moral responsibility. Detailed event-causal libertarianism positions have recently been set out by Robert Kane, Laura Ekstrom, and Mark Balaguer. A concern is that based on this picture the events preceding a decision will leave it open whether it will occur, and so whether it does is not settled by anything about the agent. Thus, whether it occurs or not is just a matter of luck, and intuitively, the agent lacks the control required to be morally responsible for the decision. This *luck objection* motivates *agent-causal libertarianism,* according to which free will of the sort required for moral responsibility is accounted for by the existence of agents who, as substances, possess a causal power to settle whether a decision is made without being causally determined to do so. To avoid the luck objection, it is essential that the causation involved in an agent's settling whether a decision is made not be a matter of causation among events but rather be an instance of the agent *as a substance* causing a choice. The agent-causal position is developed by Immanuel Kant, Thomas Reid, Richard Taylor, Roderick Chisholm, Timothy O'Connor, and Randolph Clarke. A third type of libertarianism is *noncausal* and has been proposed by Henri Bergson, Carl Ginet, Hugh McCann, and Stewart Goetz. A concern that motivates the noncausal proposal is that if our decisions are governed by any sort of causal law, they cannot be freely willed. Bergson argues that although action occurs in time, the time of conscious agency does not resolve into the kinds of quantities or magnitudes that allow them to be subject to causal laws. On Ginet's conception, the key conditions for a basic action's being free are simply that it be uncaused, that it has an agent as a subject, and that it features an "actish phenomenological" feel. Instead of the "actish" feel, McCann specifies that the action must be intentional, and intrinsically so.

In opposition to the noncausal view, critics have argued that without a causal relation between agent and action, the agent will not exercise the kind of control required for moral responsibility. Against all varieties of libertarianism, it has been objected that the sort of free will proposed cannot harmonize with our best physical theories. Our choices produce physical events in the brain and in the rest of the body, and these events are governed by physical laws, which may, for example, be deterministic. It would appear that an agent's undetermined choice could not cause an event in the body that has a deterministic causal history that traces back to a time before the agent came to exist. In response, the libertarians have attempted to make it credible that our actions could be freely willed in the sense they advocate, given the evidence we have about the physical laws.

Compatibilism

Retaining the legitimacy of our ordinary attitudes toward human actions and at the same time regarding them as causally determined has been so attractive that a large proportion of philosophers, both historical and contemporary, classify themselves as compatibilists. Two varieties of this stance might be differentiated. The first, and more common, sort aims to differentiate causal circumstances of actions that exclude moral responsibility from those that do not. The core idea is that moral responsibility requires some type of causal integration between the agent's psychology and her action, while it does not demand the absence of causal determination. This type of compatibilism is usually developed by surveying our

intuitions about blameworthiness and praiseworthiness in specific kinds of examples—involving, for instance, coercion, addiction, mental illness, hypnotism, and brainwashing. These reactions are then employed to motivate conditions on the causal integration required for moral responsibility. Important compatibilisms of this kind have been proposed by the ancient Stoics, Thomas Hobbes, John Locke, Gottfried Leibniz, and David Hume and in the past half-century by Harry Frankfurt, Gary Watson, John Martin Fischer and Mark Ravizza, Daniel Dennett, Susan Wolf, R. Jay Wallace, Ishtiyaque Haji, Alfred Mele, Dana Nelkin, and Michael McKenna.

The second kind of compatibilism, developed by P. F. Strawson, with a historical precedent in Hume, claims that the truth of determinism is irrelevant to whether we have the sort of free will required for moral responsibility. In this view, the basis of moral responsibility is to be found in reactive attitudes such as indignation, moral resentment, guilt, and gratitude. For example, the fact that certain kinds of immoral actions occasion moral resentment or indignation is what constitutes their agents being blameworthy for performing them. Moreover, justification for claims of blameworthiness and praiseworthiness ends in the system of human reactive attitudes. Because moral responsibility has this type of basis, the truth or falsity of determinism is immaterial to whether we are justified in holding agents morally responsible.

One major objection to compatibilism builds on the idea that the sort of free will at issue is the ability to do otherwise, and this ability is precluded by determinism. The *consequence argument,* crafted by Carl Ginet and Peter van Inwagen, among others, sets out this idea with rigor and precision. The crucial contention is that if determinism is true, then facts about the remote past and the laws of nature fix or entail every subsequent fact, including facts about actions, and this rules out the ability to do otherwise. Another prominent type of argument against compatibilism begins with the intuition that if an agent is causally determined to act by, for example, scientists who manipulate her brain, then she is not morally responsible for that action, even if she meets the proposed compatibilist conditions on moral responsibility. The next step is to argue that there are no differences between such manipulated agents and their ordinary deterministic counterparts that can justify the claim that the manipulated agents are not morally responsible, while the ordinary determined agents are. Such manipulation arguments have been advanced by Carl Ginet, Robert Kane, Derk Pereboom, and Alfred Mele.

Determinism

Hard determinism is espoused by Baruch Spinoza, Paul Holbach, Joseph Priestley, Ted Honderich, and Bruce Waller. Galen Strawson and Derk Pereboom argue for the related view according to which we lack the sort of free will required for moral responsibility whether determinism or indeterminism is true. Critics of such a skeptical view about free will have objected that it threatens our self-conception as deliberative agents, it renders morality incoherent, it undermines the reactive attitudes that lie at the core of human interpersonal relationships, and it jeopardizes any effective means for controlling criminal behavior. Skeptics have countered by arguing that if determinism were true, the beliefs, desires, and choices that constitute the deliberate process would be causally efficacious, and thus our self-conception as deliberative agents would remain in place; that the skeptical position would not undercut the truth of moral principles because they are independent of claims about free will; that while this conception might count attitudes such as indignation and resentment as irrational, alternative attitudes that are legitimate by its standards, such as moral concern and sorrow, could still be sufficient to sustain good interpersonal relationships; and that the skeptical view can endorse measures such as preventive detention and rehabilitation, which are sufficient for effectively dealing with criminal behavior.

Derk Pereboom

See also Agency; Causation in the Social Sciences; Causes Versus Reasons in Action Explanation; Determinism; Free Will in the Social Sciences; Laws of Nature; Libertarianism, Metaphysical

Further Readings

Balaguer, M. (2009). *Free will as an open scientific problem.* Cambridge: MIT Press.

Bergson, H. (1910). *Time and free will* (F. L. Pogson, Trans.). London, England: Allen & Unwin.

Clarke, R. (2003). *Libertarian theories of free will.* Oxford, England: Oxford University Press.

Dennett, D. (2003). *Freedom evolves.* New York, NY: Viking.
Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility.* New York, NY: Cambridge University Press.
Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy, 68,* 5–20.
Ginet, C. (1997). Freedom, responsibility, and agency. *Journal of Ethics, 1,* 85–98.
Hobbes, T. (1840). Of liberty and necessity. (A treatise, wherein all controversy concerning predestination, election, free-will, grace, merits, reprobation, etc., is fully decided and cleared, in answer to a treatise written by the Bishop of London-Derry, on the same subject). In W. Mollesworth (Ed.), *The English works of Thomas Hobbes* (Vol. 4, pp. 229–278). London, England: John Bohn. (Original work published 1654)
Holbach, P. H. T. (1835). The system of nature (H. D. Robinson, Trans.). New York, NY: G. W. & A. J. Matsell. (Original work published 1770 in French as *Le système de la nature*)
Honderich, T. (1998). *A theory of determinism.* Oxford, England: Oxford University Press.
Hume, D. (1978). *A treatise of human nature.* Oxford, England: Oxford University Press.
Inwood, B. (1985). *Ethics and human action in early stoicism.* Oxford, England: Oxford University Press.
Kane, R. (1996). *The significance of free will.* New York, NY: Oxford University Press.
Kant, I. (1996). *Critique of practical reason* (M. Gregor, Trans.). Cambridge, England: Cambridge University Press.
Lucretius. (1982). *De rerum natura* [On the nature of things] (W. H. D. Rouse, Trans.; Loeb Classical Library). Cambridge, MA: Harvard University Press.
Nelkin, D. (2011). *Making sense of freedom and responsibility.* Oxford, England: Oxford University Press.
Origen. (1913). *De principiis/Peri Archôn/*[On the principles]. In P. Koetschau (Ed.), *Origenes werke* [Works of Origen] (Vol. 5). Charleston, SC: Nabu Press.
Pereboom, D. (2001). *Living without free will.* Cambridge, England: Cambridge University Press.
Priestley, J. (1965). *A free discussion of the doctrines of materialism and philosophical necessity, in a correspondence between Dr. Price and Dr. Priestley* (Pt. 3). In J. A. Passmore (Ed.), *Priestley's writings on philosophy, science, and politics* (pp. 147–152). New York, NY: Collier. (Original work published 1788)
Reid, T. (1983). *Essays on the active powers of man.* In W. Hamilton (Ed.), *The works of Thomas Reid, D.D.* (pp. 512–530). Edinburgh, Scotland: Maclachlan & Stewart.
Scotus, J. D. (1986). *Duns Scotus on the will and morality* (A. Wolter, Ed. & Trans.). Washington, DC: Catholic University of America Press.
Spinoza, B. (1985). *Ethics.* In E. Curley (Ed. & Trans.), *The collected works of Spinoza: Vol. 1.* Princeton, NJ: Princeton University Press.
Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy, 48,* 1–25.
Suarez, F. (1965). *Disputationes metaphysicae* [Metaphysical disputations] (Vivès ed., 28 vols.). Hildesheim, Germany: Georg Olms. (Original work published 1856–1861)
Taylor, R. (1966). *Action and purpose.* Englewood Cliffs, NJ: Prentice Hall.
van Inwagen, P. (1983). *An essay on free will.* Oxford, England: Oxford University Press.

Free Will in the Social Sciences

Free will is the notion that, despite myriad causal factors that may influence human action, each one of us holds the trump card in making the final decision over our own behavior. Free will is not the claim that human action is random, which is fortunate because that would seem open to empirical refutation (and would also reduce human behavior to a simply haphazard concatenation of movements). Instead, it is a claim about the in-principle "openness" of human systems, given the virtually limitless capacity for human thought and the power that our ideas seem to have over our actions.

Some philosophers have held the view that the problem of human free will—second only perhaps to the problem of consciousness—is the most difficult one in philosophy. Indeed, one thing that makes it so difficult is that free will seems to be an irreducibly philosophical matter, in that there appear to be no obvious experiments that social scientists could perform to settle the matter. Consequently, we are left with the difficulty that we have ample subjective evidence for our feeling of freedom, but none of it is scientifically measurable or even shareable with other people. This puts us in a situation where each of us is all but certain that we have free will, but when the time comes to prove it, we are stuck with nothing more than our internal feelings, which have shown themselves to be unreliable in other venues.

In the absence of hard evidence, and the presence of strong feelings, we are left to argue.

One of the main arguments that philosophers have had concerning free will is whether it is compatible with determinism. *Incompatibilists* believe that it is not; that the truth of determinism would obviously undermine the notion of human choice—and with it the idea that humans are morally responsible for their actions—because it degrades the causal efficacy of intentions that is necessary for both freedom and responsibility. This conclusion is disputed by *compatibilists,* who maintain that the truth of determinism would not necessarily undermine human freedom (or moral responsibility) by pointing to several thought experiments where our notion of free choice would seem to survive the idea that our behavior was completely predictable or even fully determined (e.g., if there were a Deity). A third view—sometimes called *libertarianism* or the *Theory of Agency*—argues that compatibilist accounts are misguided but that, fortunately, this makes no difference since determinism is not true. The philosophical argument required to demonstrate this claim is difficult and seems to hang on the notion that the causal power of human action is unique and can be located squarely within human decisions. Critics of this view, however, have maintained that this account is simply tantamount to saying that humans have free will, and thus, this brings us back to the beginning of the debate.

In light of the notorious failure of philosophers to make progress in solving these questions, it is tempting to revisit the possibility that one can gather empirical evidence that is relevant to the question of free will. Indeed some neuroscientists have been searching for a way to test the subjective hypothesis of free will by looking past the social sciences and focusing directly on the brain.

In one such experiment in 1983, Benjamin Libet famously discovered that brain activity correlating with a person's decision to move his or her finger can be detected 300 milliseconds *before* the person's conscious decision to act. This finding is widely taken by some to threaten the notion of free will, in that it purports to show that unconscious brain processes can commit us to a certain course of action even before we become aware of it. This study has been challenged, however, for relying too heavily on a subject's self-report of when he or she formulated the intention to act, as well as because of the obvious problem that it would take a certain amount of time for a subject not only to *decide* to act but also to decide to *report* that he or she had decided to act, which may account for the interval. Follow-up studies, which have eschewed subjective reports in favor of direct examination of the brain using functional magnetic resonance imaging machines, have deflated such criticisms while providing evidence that corroborates the direction of the original effect.

Even as experimental evidence mounts in favor of the existence of an interval between brain activity and conscious intention, some commentators have been reluctant to draw any ominous conclusions for the notion of free will, arguing instead that what we need is to revise our views of the mechanism by which human actions take place. Perhaps human decision making is best understood not as a conscious intention followed by an overt action but instead by the subjective feeling that occurs when the brain transforms an unconscious plan into action. Others have said that perhaps human decision making is merely the occasion upon which both our intentions and our actions arise as collateral effects of an as yet unknown brain process. Whether such reconceptions of intentionality (and causality) would allow for the survival of traditional notions of human free will remains a matter of ongoing debate among neuroscientists and philosophers.

Thus, whether human action is truly free or whether we are merely manipulated into thinking so by some causal mechanism that guides both our thoughts and our actions is one of the great unanswered questions of contemporary philosophy.

Lee McIntyre

See also Causes Versus Reasons in Action Explanation; Determinism; Explanation Versus Understanding; Free Will, Philosophical Conceptions of; Intention, Social Psychology of; Libertarianism, Metaphysical; Unconscious Social Behavior

Further Readings

Churchland, P. (2002). *Brain-wise: Studies in neurophilosophy.* Cambridge: MIT Press.

Harris, S. (2011). *The moral landscape: How science can determine human values.* New York, NY: Bantam Books.

Watson, G. (Ed.). (1983). *Free will.* New York, NY: Oxford University Press.

Game-Theoretic Modeling

This entry exemplifies the process by which models of strategic behavior are constructed (requiring a modicum of mathematics for a full appreciation). Strategic behavior arises when the outcome of an individual's actions depends on actions taken by other individuals. For example, whether it is advantageous for drivers negotiating a four-way junction to assume right-of-way depends on whether other drivers concede the right-of-way. Likewise, whether one prospers from moral behavior depends on whether others do the right thing. If an interaction among individuals gives rise to strategic behavior and can be described mathematically, then we call this description a game and each individual a player. Thus, a game is a model of strategic interaction, and game-theoretic modeling is the process by which such games are constructed.

A game has four key ingredients. First, there are at least two players, who may be either specific actors or individuals drawn randomly from a large population: Drivers at a four-way junction may be either neighbors or strangers. Correspondingly, the game is either a community game or a population game. Second, each player has a set of feasible plans of action—or strategies—which are constrained by the information structure of the interaction. For example, drivers at a four-way junction can condition their behavior on their lateness relative to others only if they are sufficiently aware of it. Let the lateness of two such drivers who wish to turn left simultaneously be represented by the random variables X and Y, taking values between 0 (*unbelievably early*) and 1 (*desperately late*); the sample space for their joint distribution is the unit square $0 \le x, y \le 1$. Then it is possible for the first driver, or Player 1, to play a strategy u, defined by "Go if $X > u$; wait if $X \le u$," while the other driver, or Player 2, plays strategy v, "Go if $Y > v$; wait if $Y \le v$." We call this game "Crossroads."

Third, there is a well-defined reward for each player from any potential strategy combination. In Crossroads, let τ be the time it takes a driver to traverse the junction, and suppose that the drivers discount this delay by a fraction η of their earliness. Thus, if $X \le u$ and $Y > v$, then the delay of τ as Player 2 traverses the junction is experienced as $-\tau\{1 - \eta(1 - X)\}$ by Player 1, whereas if $X > u$ and $Y \le v$, then the delay of τ is experienced as $-\tau\{1 - \eta(1 - Y)\}$ by Player 2 but is 0 for Player 1. If both drivers either "Go" or "Wait" in the first instance, then there is an additional delay of δ or $\in$, respectively, with $\delta > \in$, as they sort out who will subsequently drive away first. Assuming that it is equally likely to be either driver and that delays should be as short as possible, so that the negatives of delays serve as payoffs, Player 1's payoff is the random variable

$$F_1(X, Y, u, v) = \begin{cases} -\tau\{1 - \eta(1 - X)\} & \text{if } X \le u,\ Y > v \\ -(\delta + \tfrac{1}{2}\tau\{1 - \eta(1 - X)\}) & \text{if } X > u,\ Y > v \\ -(\in + \tfrac{1}{2}\tau\{1 - \eta(1 - X)\} & \text{if } X \le u,\ Y \le v \\ 0 & \text{if } X > u,\ Y \le v \end{cases}$$

Player 1's reward from the strategy combination (u, v) is the expected value of F_1, which we denote by $f_1(u, v)$. That is,

$$f_1(u, v) = \int_0^1 \int_0^1 F_1(x, y, u, v)\, g(x)\, g(y)\, dx\, dy,$$

where g is the probability density function of X and Y's common distribution. By symmetry, Player 2's reward is $f_2(u, v) = f_1(v, u)$.

The last main ingredient of any game is a solution concept. An appropriate one for community games is the Nash equilibrium, a strategy combination from which no individual has a unilateral incentive to depart. Equivalently, a Nash equilibrium is a combination of mutual best replies. In Crossroads, if lateness has a uniform distribution, or $g(x) = g(y) = 1$, and drivers are fast enough to ensure that $\frac{1}{2}\tau < \delta$, then $f_1(u, v)$ is maximized for a given v by

$$u = \begin{cases} 1 & \text{if } 0 \le v \le 1 - \theta \\ \dfrac{1 - \theta + \lambda - v}{\lambda} & \text{if } 1 - \theta < v < 1 - \theta + \lambda, \\ 0 & \text{if } 1 - \theta + \lambda \le v \le 1 \end{cases}$$

where $\theta = (\in + \frac{1}{2}\tau)/(\in + \delta)$ and $\lambda = \frac{1}{2}\eta\tau/(\in + \delta)$. Because $\frac{1}{2}\tau < \delta$ implies $\lambda < \theta < 1$, 0 or 1 is the unique best reply to 1 or 0, respectively, and $v^* = 1 - \theta/(1 + \lambda)$ is the unique best reply to itself. Thus, by symmetry, there are three Nash equilibria, namely, (0, 1), (v^*, v^*), and (1, 0). An appropriate solution concept for population games is John Maynard Smith's evolutionarily stable strategy, or ESS. Strategy v is an ESS if, when adopted by a population, it does not pay anyone to switch from v to any other strategy. For Crossroads, v^* is the unique ESS.

Space constraints preclude a discussion of the numerous subtleties that arise in connection with game-theoretic modeling. For example, a theoretical distinction is often observed between noncooperative games, in which any agreements are self-enforcing, and cooperative games, in which all agreements are externally enforced and the Nash equilibrium is superseded by various cooperative solution concepts. The distinction is not always useful, however, especially in games of coalition formation. These and other subtleties are discussed at length in the suggested further readings.

In philosophy and the social sciences, game-theoretic modeling is valuable because it enables one to explore the logic of a verbal argument rigorously. For example, in a population of drivers at the ESS of Crossroads, the average delay is $-f_1(v^*, v^*)$ and is readily shown to decrease with η. Thus, drivers who discount experience shorter average delays than drivers who do not discount; and the more they discount, the less they wait. Although this result may be highly plausible on the basis of a purely verbal argument, only a game-theoretic model can convincingly nail it down.

Mike Mesterton-Gibbons

See also Conventions, Logic of; Cooperation, Cultural Evolution of; Cooperation/Coordination; Equilibrium in Economics and Game Theory; Evolutionary Game Theory and Sociality; Mathematical Models, Use in the Social Sciences

Further Readings

Gintis, H. (2009). *Game theory evolving: A problem-centered introduction to modeling strategic interaction* (2nd ed.). Princeton, NJ: Princeton University Press.

Levin, S. A. (Ed.). (2009). *Games, groups, and the global good*. Berlin, Germany: Springer.

Mesterton-Gibbons, M. (2001). *An introduction to game-theoretic modeling* (2nd ed.). Providence, RI: American Mathematical Society.

Mesterton-Gibbons, M., Gavrilets, S., Gravner, J., & Akçay, E. (2011). Models of coalition or alliance formation. *Journal of Theoretical Biology, 274,* 187–204.

GENEALOGY

Genealogy is an approach for the study of formative historical events, pioneered by Friedrich Nietzsche and developed further by Michel Foucault. There are also important links between genealogy and comparative historical sociology, visible in the works of Max Weber and some of his followers. The genealogy of a particular political institution or social practice is concerned with the exact manner in which it emerged and the lasting effects it might exert even after the institution or practice ceased to exist.

Nietzsche

In philosophy and in the social sciences, the term *genealogy* is closely associated with the work of Friedrich Nietzsche, in particular *On the Genealogy*

of Morality and *The Birth of Tragedy.* These are Nietzsche's only books that are sustained philosophical investigations, not aphorisms, showing the significance of genealogy for Nietzsche. The central underlying idea is that contingent historical events might play a decisive role in the very formation of institutions or even personality structures that are taken for granted, or considered as "natural" or "rational." What looks like a systematic whole is therefore a sum of historical layers, which only become saturated with rationality because of the durability of components.

The clearly combative intent is revealed by the subtitle of the *Genealogy,* often omitted in English translations: "A Polemical Tract." It also transpires from the word *genealogy* itself, associated with aspects of the *ancien régime* most unacceptable to the modern mind: the long lineages of patriarchs, culminating in the genealogy of Jesus in the house of David and the similar family ancestry of feudal aristocracy.

Genealogy, as championed by Nietzsche, implies two specific claims about studying a concrete phenomenon, be it morality, modernity, or a political system. The first, indicated by the word *birth,* is that the manner in which a particular entity comes into being matters. Thus, political institutions or social practices coming into being at a certain moment are concrete and specific responses that gather their "justification" over time. The second is concerned with lasting effects: Even once a particular institution ceased to function, its ways of acting and thinking might be carried over for a considerable time.

Questions of birth or emergence can be considered as relevant in two different manners, corresponding to Darwinian or Lamarckian approaches about heredity. In one case, something positive is transmitted, whether as genes, modes of behavior, or a certain "spirit"; in the other, the particular conditions characterizing a "birth" will be carried over. Here, there is a significant displacement of emphasis in between Nietzsche's two major relevant books. This is visible in dropping the original subtitle of *Birth of Tragedy,* "Out of the Spirit of Music," in the second edition published just before the *Genealogy,* and in the preface to the *Genealogy,* best epitomizing Nietzsche's new focus on conditions of emergence.

Still, in contrast to deconstruction in the style of Jacques Derrida, this does not mean that for Nietzsche genealogy was a purely negative undertaking. Continuity, or the ability to persist, had a value for Nietzsche, just as he was interested in the rebirth of what lay dormant (see the Renaissance). This animated his interest in the cultivation and promotion of human qualities, as was visible in his concern with questions about "what is noble" and who is the "good European." Genealogy for Nietzsche also had a self-reflexive component: It was an attempt at a self-understanding.

Foucault and Weber

Michel Foucault explicitly returned to Nietzsche's genealogy, using "birth" in titles or subtitles and being credited with developing a "genealogical method." He defined the middle part of his work as "genealogy of power," acknowledging the second essay of the *Genealogy* as the inspiration behind *Discipline and Punish*. The period is marked, at both ends, by crucial methodological statements: the 1971 essay "Nietzsche, Genealogy, History," a meticulous study of Nietzsche's related writings; and a 1978 roundtable discussion, which focuses on the significance of events for historical understanding. Foucault's final statement about genealogy as method is in the introduction to *The Use of Pleasure.*

Foucault's works had a great resonance over broad areas in the social sciences. However, Nietzsche's approach had a major impact on historically oriented social and political theorists decades before. The central figure is Max Weber, whose comparative sociology of religions and civilizations followed a genealogical design. This is visible in *The Protestant Ethic and the Spirit of Capitalism,* which claims that the moving spirit of capitalism is rooted in the inner-worldly turn of medieval monastic asceticism (the third essay of the *Genealogy* was devoted to the "ascetic ideal"); the first page of the section on religion in *Economy and Society,* focusing on the conditions and effects of social actions; and the importance attributed to "stamping experiences" in the introduction to the *Economic Ethic of World Religions*. Such a Nietzschean inspiration characterizes those who followed the spirit of Weber's work: Norbert Elias, with his sociogenesis and psychogenesis; Eric Voegelin, who argued that the spirit of the modern nation-state, characterized as "intramundane eschatology," grew out of the apocalyptic expectations of medieval sects and who introduced the term *historiogenesis;* and Reinhart Koselleck, with his "pathogenesis" of modernity.

The genealogical approach has its significance for philosophy as well, as it tries to move beyond Hegelian dialectics and the phenomenology of Husserl, efforts to correct the ahistoric formalism and rationalism of Kant. Hegel attempted to render Kantian categories dynamic and historical, but his dialectic was entrapped in the same dichotomizing and dualistic thinking. In contrast to this, Nietzsche's genealogy emphasized the "middle," with polarization into extremes being a possible outcome, not a precondition for understanding. Husserl wanted to go back to "the things themselves," beyond Kantian and Hegelian abstractions, but the absence of a historical focus compromised concreteness. Genealogy combines a historical analysis of categories of thought with a study of formative events.

Problems With Genealogy

Nietzsche-inspired genealogy faces two main problems. The first concerns meaning and value: A genealogy that is purely deconstructive, ironically showing that our cherished values are historical constructs, can be charged with nihilism. In contrast to this, there is a "positive" aspect of genealogy, of which Foucault's late work about care of the self and truth telling is a good example. Second, neither Nietzsche nor his followers managed to specify the manner in which conditions of emergence might leave a "stamp." Concepts developed by anthropologists, like *liminality* (Arnold van Gennep and Victor Turner) or *schismogenesis* (Gregory Bateson), offer a solution. Such a perspective also illuminates the similarities between Nietzsche's genealogy and William Dilthey's philosophical anthropology as a "critique of historical reason."

Arpad Szakolczai

See also Foucault's Thought; Nihilism; Norbert Elias: Process of Civilization and Theory of Sciences; Philosophy of Sociology, History of; Weber and Social Science: Methodological Precepts

Further Readings

Elias, N. (2000). *The civilizing process: Sociogenetic and psychogenetic investigations.* Oxford, England: Blackwell. (New edition: Elias, N. (2012). *On the process of civilization.* Dublin, Ireland: University College Dublin Press)

Foucault, M. (1975). *The birth of the clinic.* New York, NY: Vintage.

Foucault, M. (1979). *Discipline and punish: On the birth of the prison.* New York, NY: Vintage.

Foucault, M. (1981). Questions of method. *I&C, 8,* 3–14.

Foucault, M. (1984). Nietzsche, genealogy, history. In P. Rabinow (Ed.), *The Foucault reader* (pp. 76–100). New York, NY: Pantheon.

Horvath, A., & Thomassen, B. (2008). Mimetic errors in liminal schismogenesis: On the political anthropology of the trickster. *International Political Anthropology, 1,* 3–24.

Koselleck, R. (1988). *Critique and crisis: Enlightenment and the pathogenesis of modern society.* Oxford, England: Berg.

Nietzsche, F. (1994). *On the genealogy of morality.* Cambridge, England: Cambridge University Press.

Nietzsche, F. (1999). *The birth of tragedy and other writings.* Cambridge, England: Cambridge University Press.

Szakolczai, A. (1998). *Max Weber and Michel Foucault: Parallel life-works.* London, England: Routledge.

Weber, M. (2002). *The Protestant ethic and the spirit of capitalism.* Los Angeles, CA: Roxbury.

Genetic Indeterminism of Social Action

After decades of discussions over whether social action may be "caused" by biological/genetic or other factors, newer research shows that any human behavior and most human traits are shaped by a dynamic interaction of diverse factors. Most of them cannot be neatly pinned down to the realm of genetics, biology, or the social environment. This entry will provide a brief overview of the development of genetic and other research on human traits and behavior and show why it is now well established that social action is a complex phenomenon that is far from being determined by any one factor.

Studying the Role of "Nature": The Twin Study Design

Studies with twins have been seen as a particularly fruitful mode to learn about the heritability of human traits. Heritability is defined as the proportion of variation in a trait in a population that can be explained by genetic factors. Twin research designs compare identical twins, who share virtually

100% of their genetic material (DNA), with nonidentical (fraternal) twins, who share about half, just like regular siblings. For example, if a study shows that identical twins *always* have the same eye color while the eye color of nonidentical twins is different in half of all cases, then the conclusion is drawn that eye color is a strongly heritable trait. The twin study design does not, and cannot, however, tell us anything about which or how many genes and mechanisms are involved in passing on heritable traits. Moreover, unlike eye color—a trait that is almost entirely heritable—most traits, such as predispositions to common diseases, certain aspects of our personality, or physical attributes such as height, are neither entirely heritable nor entirely environmental (whereby *environment* means "everything that is not genetic," including lifestyle factors, social environments, toxin exposures, etc.). Thus, in most contexts, twin studies can give us only a rough indication of the extent to which genetic and environmental factors may be involved in shaping a trait.

From Nature Versus Nurture to Gene–Environment Interaction

For a long time, social scientists, biologists, psychologists, and scholars in other disciplines argued about what factor was primarily responsible for shaping human traits in general and behavior in particular: "nature," understood as the allegedly inalienable genetic dimension of humans, or "nurture," comprising child rearing, education, and other dimensions of our social environment. Apart from racist or conservative hardliners who have a vested interest in arguing that certain characteristics are hardwired into particular groups of people, hardly anyone discusses human behavior in terms of nature *or* nurture anymore. Now, research and discussions focus on how nature and nurture interact. Not only has research in many academic disciplines shown that hardly any trait is shaped entirely by either genetic or social factors, but recent advances in research have discovered that environmental and genetic factors shape each other: They interact in complex ways that are not necessarily additive (their combined effects may be different from the sum of their individual effects). *Epigenetics* has been particularly informative in this regard: It is a multidisciplinary approach to understanding why some environmental influences, such as the consumption of particular foods, can change the way genes are chemically switched on or off. These changes to how genes are regulated (switched on and off) can be passed on to the next generation, yet without changing the DNA sequence as such.

Recently, critiques of sociobiology, such as that of Richard Lewontin, have revisited the environment–gene dichotomy, or the "inside–outside" divide, offering more subtle analyses of reciprocal interaction and intercausal change between the two relata.

What Social Sciences and Philosophy Can Contribute

Current models of behaviors and disease predispositions posit that genes and environments interact in a dynamic and potentially intergenerational manner. Yet the term *environment,* as mentioned earlier, is understood widely, including the molecular or biological environment of genes and organs, "lifestyle" and other behaviors, and the wider physical and social environments. Biological scientists may see environments as more difficult to conceptualize and operationalize than genes. This means, according to critical social scientists and philosophers, that even in contemporary research, genetic factors still tend to be prioritized or seen as the starting point, while environments external to the body tend to be poorly specified. For example, as the cultural anthropologist Margaret Lock argues, recent research on the causes of Alzheimer's disease has focused on variations at the molecular level (difference in DNA, proteins, and other cell components), sometimes in conjunction with a small number of—inadequately conceptualized—environmental influences such as gender, ethnic background, and level of education. This has come at the cost of more fine-grained investigations of the ways in which human relationships across the life span may affect gene regulation and developments in the brain.

To the "Geneticization" of the Social (and Back Again?)

Current work on genetic and environmental factors involved in shaping human action therefore highlights a place for more detailed and sophisticated ways of conceptualizing and incorporating social dimensions into epidemiologic and basic research. Beyond this, the social sciences and philosophy have provided a number of important concepts for understanding the wider significance of genes for society.

The concept of *geneticization,* for example, encapsulates the idea that genetics has increasingly become the dominant lens through which human health and behavior are understood. This is seen to have a number of negative repercussions: for example, that genetic research has attracted an unjustifiably large proportion of intellectual and financial resources, detracting from social and environmental research and initiatives.

The concept of *biosociality* turns the idea that biology determines behaviors on its head, to suggest that human interventions will, in the end, shape biology. Biosociality also implies that people are increasingly forming groups on the basis of shared biological or genetic characteristics, for example, being at risk for a particular disease, so as to influence research and ultimately the health of those with whom they share a genetic risk. Here, then, social interaction aims to determine biology.

Far from genes being immutable, and determining social action, we have learned that the effects of genes and the environment are interwoven and *both* may be seen as malleable.

Barbara Prainsack and Kate Weiner

See also Biology and the Social Sciences; Determinism; Evolutionary Ethics; Evolutionary Psychology; Intelligence; Naturalism in Social Science; Neuroethics; Neuroscience and Politics; Sociobiology

Further Readings

Gibbon, S., & Novas, C. (Eds.). (2008). *Biosocialities, identity and the social sciences.* London, England: Routledge.

Kaldis, B. (2011). Nature and society. In G. Delanty & S. Turner (Eds.), *Routledge international handbook of contemporary social and political theory* (pp. 312–327). London, England: Routledge.

Keller, E. F. (2010). *The mirage of a space between nature and nurture.* Durham, NC: Duke University Press.

Lewontin, R. (2000). *The triple helix: Gene, organism, and environment.* Cambridge, MA: Harvard University Press.

Lock, M. (2005). Eclipse of the gene and the return to divination. *Current Anthropology, 46*(Suppl.), S47–S60.

Nowotny, H., & Testa, G. (2011). *Naked genes: Reinventing the human in the molecular age.* Cambridge: MIT Press.

Prainsack, B., & Hashiloni-Dolev, Y. (2009). Religion and nationhood: Collective identities and the new genetics. In P. Atkinson, P. Glaser, & M. Lock (Eds.), *Handbook of genetics and society: Mapping the new genomic era* (pp. 404–421). London, England: Routledge.

Shostak, S., & Freese, J. (2010). Gene-environment interaction and medical sociology. In C. E. Bird, P. Conrad, A. M. Freemont, & S. Timmermans (Eds.), *Handbook of medical sociology* (6th ed., pp. 438–514). Nashville, TN: Vanderbilt University Press.

Weiner, K., & Martin, P. (2008). A genetic future for heart disease? *Sociology of Health & Illness, 30*(3), 380–395.

Gestalt Psychology

This entry presents the Gestalt school of psychology, which, among other things, in developing theories of experience and, in particular, theories of perception, provided not only the famous Gestalt-switch notion employed by Kuhn and the Wittgensteinians to assess either a certain view of radical theory change in science or the seeing-as-conception of understanding but also the general idea of wholes that was found congenial by holist views in the social sciences. The part–whole relations studied by Gestalt psychology, especially in visual or acoustic perception, offered support to specific ways of establishing anti-individualist arguments in the philosophy of the social sciences. In particular, they emphasized "functional wholes," "dynamic events," and "fields" in experience, which were also true in physics.

A usual mistake is to portray the Gestalt holist stance as espousing the motto "The whole is more than the sum of its parts." This is not actually what they maintained, since they had a more sophisticated understanding of whole–part relations. As one of the founders of Gestalt psychology, Kurt Koffka (1935/2001), put it, "It is more correct to say the whole is something else than the sum of its parts, because summing is a meaningless procedure, whereas the whole–part relationship is meaningful" (p. 176).

Definition of Gestalt

Gestalt psychologists emphasized the study of consciousness, but not in the same way as structuralist psychologists did, who used controlled introspection in sensory experiments. Unlike atomist conceptions of psychology, which searched for the elements

of experience regarded as primary, Gestalt theory emphasized that both perceived forms (*Gestalten*) and the experienced world are already structured wholes: Experience involves "sensory organization" in which the parts (of experience) are dependent on the whole. Gestalt theorists rejected the view of structuralists and associationists that consciousness consists of associated elements. For example, the reversible figure of the Rubin vase, which also appears as two faces, is fundamentally different from collections of sensations—that is, a mere aggregate of sense-data. Later Gestaltists accepted behavioral wholes, such as the demonstration of an ape that uses insight to put together two bamboo sticks to reach a banana, instead of learning gradually through trial and error. Though they emphasized consciousness, they linked it with the brain by assuming isomorphism.

In terms of methodology, classical Gestalt psychology depended heavily upon the single demonstration or *experimentum crucis;* however, by the mid-20th century, Gestalt social psychology had bridged experiment and social action.

The Austrian psychologist and philosopher Christian von Ehrenfels, who fist coined the term *Gestalt* in the special sense to be adopted by Gestalt psychologists, spoke of *Gestaltqualitäten*. Going beyond the definition of Gestalt as mere shape or form, he emphasized the presence of such qualities in all experience by pointing out that no sensation arises by local stimuli alone but, rather, in relation to sets of stimuli or in a "togetherness" with a background environment or perceptual context, thus pioneering the notion of "whole" in experience.

Prehistory of Gestalt in Philosophy and Medicine

Although Gestalt psychology adhered to scientific naturalism and experiment, it had deep philosophical roots. Philosophers following Kant began to distinguish the act of judging from the judged object. The proper object for idealist philosophers was not a mental picture but the entire object or situation. "Cheetahs exist" or "It's raining" are unitary judgments of a "state of affairs" (*Sachverhalt*), also translated as "atomic facts," which are not combinations of ideas in a proposition. The philosophers Hermann Lotze, Carl Stumpf, Alexius Meinong, Kasimir Twardowski, Franz Brentano, Edmund Husserl, and Ludwig Wittgenstein developed this critique of judgment. In Berlin before and after World War I, Carl Stumpf agreed with William James that we experience a "stream of consciousness" from which we distinguish "mental functions." In visual space, we perceive location directly, not through muscle movements in the retina. Two tones sounded together yield a third tone different from the component tones.

In 19th-century perception research, Ewald Hering challenged Hermann Helmholtz's description of the eye as a camera in which we bring together sensations by unconscious inference. A zigzag hole in a white cardboard can appear as a black patch or as a hole. Each perception has psychological reality as a "seen object," even if the retinal image is the same. Christian von Ehrenfels noted that one can play a melody in one key and then in another key and immediately recognize the tune. Similarly, Austrian philosophers in Vienna and Graz contributed a model of cognition based on the Müller-Lyer illusion, in which the length of two lines is judged longer or shorter based on illusions created by arrows at their ends.

In the early 20th century, Gestalt psychologists experimented with tone color, figural aftereffects, illusions, and perceptual constancies. Brain scientists explored equipotentiality and mass action, behavioral neurology, and self-actualization. Social psychologists extended Gestalt theory to social *fields* of force in a life-world and to humanizing the workplace. Gestalt therapists encouraged healthy contacts with the world through reorganization of the self in a social field.

Gestalt in Human Perception and Animal Insight Experiments

Stumpf's students in Berlin, Max Wertheimer, Kurt Koffka, and Wolfgang Köhler named and popularized Gestalt psychology elsewhere, principally in the United States, based upon a wide range of perceptual, learning, and cognitive experiments. Wertheimer's classic phi phenomenon experiment used slits of light in a revolving tachistoscope wheel to produce a perception of apparent motion. Exposure intervals, as well as the color, arrangement, and size of the stimuli, could be varied.

Kurt Koffka introduced the "white tablecloth experiments," where a subject encounters a white and a black tablecloth, but the black one has greater illumination. Subjects have no difficulty identifying that the white is white, even though the "proximal stimuli" of the black one are brighter. The subject utilizes a "color gradient" to make a Gestalt perceptual judgment. This involved a new understanding of stimulus. A hungry fish bites the worm, but the satiated one does not; the same stimulus object may elicit a different pattern of response as a "direct experience correlate of the stimuli." No hypothetical psychological processes are involved. In his book *The Growth of the Mind,* Koffka applied the anthropologist Lucien Lévy-Bruhl's concept of the primitive mind to the child, maintaining that the child's mind grows by differentiation, from the whole to the parts. In Russia, the developmental psychologists Lev Vygotsky and Alexander Luria drew on the work of Koffka, Kurt Lewin, Tamara Dembo, and Anita Karsten, bringing a Gestalt perspective into their cultural-historical school.

In the Tenerife Island experiments with apes during World War I, Wolfgang Köhler observed apes overcoming obstacles to reach a goal object—by stacking boxes. He and others (Carl Bühler, Charlotte Bühler) argued against Edward Lee Thorndike's learning experiments with cats, in which a cat engaged in random efforts to get out of a puzzle box and the experimenter graphed the time taken to do so by a learning curve. Such instrumental learning abstracted from the animal's behavior, whereas Kohler's insight phenomenon has face validity.

Also in Germany, Kurt Lewin treated actions as "behavioral wholes" in which emotion and volition contribute to a specific situation with an end product. Bluma Zeigarnik and Maria Rickers-Ovsiankina showed that memory is better for uncompleted tasks. Tamara Dembo manipulated the experiment–subject relationship by causing frustration in subjects trying to complete an assigned task.

Gestalt Social Psychology

Dembo's observation of "firmness" and "loosening" of tension systems led to Lewin's later concepts of psychological "life space" and "social fields of force." Evgenia Hanfmann and Rickers-Ovsiankina took Lewin's program into counseling, giving attention to the orientation of students to the clinical situation. The work of many of Lewin's students contributed to social action research by the Society for the Psychological Study of Social Issues in the United States.

Lewin also focused on how the situation appears to the actor, and he diagrammed the needs of the person in different layers. He included motivation forces, as well as self and field, but opposed historical explanation by childhood experiences (psychoanalysis) or past learning (behaviorism). Physics provided the concepts of *field of force* and *vectors* and ahistorical laws by which social Gestaltists described personality and *social fields.*

The experimenter first became a subject of psychological attention in Lewin's research. By contrast, contemporary research about character types and mental traits is abstracted from groups of persons, and the experimenter is invisible. Lewin's largely East European female research group had a cosmopolitan awareness of social-psychological questions, recognizing that personality is embedded in social situations.

Another émigré, Fritz Heider, proposed balance theory: If A dislikes B and B dislikes C, then he predicted that A would like C. Or if a person likes President Eisenhower and he was reported to have done something bad, then one could protect equilibrium by disbelieving this report. Taking this balance theory further, Leon Festinger found that if students were offered a small reward for giving a speech that contradicts their beliefs, they are more likely to change their attitude than if they were given a larger reward. He called this "cognitive dissonance." Heider and Festinger did not share Lewin's emphasis on the Gestalt social field; theirs was an individual social psychology.

In medicine, the neurologists Jakob von Uexküll, Konstantin von Monakow, and Kurt Goldstein emphasized "holism" in brain science. Some cultural psychologists defended a "racializing" national socialism with ideas of "racial soul," "purity," and "folkish study of humanity." Others built early careers by publishing in Nazi journals and weaving in statements of party orthodoxy.

Since the mid 20th century, many personality and social psychologists have argued that the historical approach based on frequencies can be joined with social field accounts employing experimenter–subject interaction. David Krech and R. S. Crutchfield, in their 1948 textbook, included historical frequency data along with group phenomena; however, the emphasis was on the psychological states

of individuals. By comparison, when Solomon Asch performed multiple trials with confederates to produce a conformity effect in the subject's judgment of the length of a line, the emphasis became the individual and the group: They mutually influence one another in socially engaged psychological behaviors. Muzafer Sherif and Carolyn Wood Sherif manipulated groups of boys at summer camp to become friends, then to compete as rivals, and finally to reconcile through working toward a superordinate goal. Their concern with the origin and development of norms marks them too as truly social psychologists, the direction in which Lewin had pointed.

Gestalt Therapy

A vibrant clinical branch of Gestalt therapy grew out of the work of Fritz Perls and Laura Posner Perls (who studied color contrast with the Gestaltist Adhémar Gelb in Frankfurt and cowrote *Ego, Hunger and Aggression* in 1947). Paul Goodman then wrote *Gestalt Therapy* with Fritz Perls and Ralph Hefferline in 1951. They emphasized the client's "resistance" or bodily armor, drawing from their mentor Wilhelm Reich and from Jacob L. Moreno's psychodrama. In doubling or "hot seat," clients kneel to symbolize the hidden aspect in their lives while saying what they would feel but never tell.

Since the 1950s, Gestalt therapy schools have arisen across North America and the world. The psychoanalytic concept of "mechanisms of defense" (Anna Freud) evolved into the restrospective "resistances to contact" (Perls). For example, clients may be asked to pay attention to bodily sensations while walking in a circle. Erving and Miriam Polster changed the emphasis to "working with the resistance." A therapist should look at the person the way you would view a sunset, with love and affection. One technique is mining for stories. A client dreamed that she and her sister had escaped together. From whom, asked the therapist. From people. From what people? From our parents. By insisting upon concreteness, the therapist facilitates the client's transition from A to B. Then, focused emotional work becomes possible.

Relativism Versus Realism: Philosophy of Science Versus Gestalt Psychology

Philosophers of science debated the significance of the so-called Gestalt switch in terms of realism versus relativism. This raises issues of the foundations of scientific truth when (Kuhnian) paradigms change.

Wittgenstein, in his *Philosophical Investigations* (1952), introduced the figure of a duck-rabbit. It can be involuntarily seen as one, then the other, but never as both simultaneously. Here, the same sense-data (lines or figures perceived) give rise to two different perceptions (either duck or rabbit)—in other words, to one precept at a time. So, the argument goes, the underlying reality or fact of the matter may be one, but it is perceived differently given the diverse theories affecting our observation—hence the relativism in scientific theory formation. Transposed to how a paradigm switch (as a Gestalt switch) works in scientific theory replacement, we can understand the difference, for example, between how Tycho Brahe (a geocentrist) saw the sun rising and Kepler (a heliocentrist) saw the earth's revolutions lowering it with respect to the sun. While the underlying reality is supposed to be one, the theoretical lenses through which it is viewed are different. In the hands of N. R. Hanson and Thomas Kuhn, a paradigm switch worked like a Gestalt switch.

Wittgenstein explained this by sensory and intellectual contributions, a legacy perhaps of his predecessors in the Austrian school of philosophy, including Franz Brentano and Edmund Husserl. In another example, three dots are seen as a triangle; the dots are the sensory portion, and the triangle is the judgment.

Gestalt psychologists argue that this returns us to the intellectualism of the tradition they rejected. They are realists in asserting that perception itself is structured. Reality is given to us with a rich phenomenology. The conditions that produce this Gestalt switch depend on the "definiteness" of stimulus conditions; if they are poor, switching may occur. Such conditions occur in everyday life, yet we perceive accurately.

The implications for philosophy of science were profound. If the Gestalt switch consists of both sensory and intellectual entities (the underlying facts plus the theoretical lenses), they must be connected by associations. How can scientists then communicate about the facts if they simply interpret according to their experience? Thomas Kuhn's use of the duck-rabbit to illustrate a scientific paradigm shift thus invokes an older epistemology, one that Gestalt psychologists argued against with their perceptual experiments. He represents the relativism that has emerged in the philosophy of science since the mid 20th century; the early-20th-century Gestaltists remain perceptual realists.

William R. Woodward

See also Emergence; Hermeneutics, Phenomenology, and Meaning; Holism, in the Social Sciences; Observation and Theory-Ladenness; Phenomenological Schools of Psychology; Philosophical Psychology, History of; Systems Theory

Further Readings

Asch, M. G. (1996). *Gestalt psychology in German culture: Holism and the quest for objectivity.* New York, NY: Cambridge University Press.

De Rivera, J. (1976). *Field theory as human science. Contributions of Lewin's Berlin group*. Hillsdale, NJ: Lawrence Erlbaum.

Ellis, W. D. (Ed.). (1938). *A source book of Gestalt psychology.* London, England: Routledge & Kegan Paul. (Contains five chapters by M. Wertheimer, three by Köhler)

Festinger, L. (1957). *A theory of cognitive dissonance.* Evanston, IL: Row, Peterson.

Koffka, K. (1924). *The growth of the mind: An introduction to child psychology* (R. M. Ogden, Trans.). New York, NY: Harcourt, Brace.

Koffka, K. (2001). *Principles of Gestalt psychology.* London, England: Routledge. (Original work published 1935)

Köhler, W. (1969). *The task of Gestalt psychology.* Princeton, NJ: Princeton University Press.

Köhler, W. (1992). *Gestalt psychology: An introduction to new concepts in modern psychology.* New York, NY: Liveright. (Original work published 1947)

Kuhn, T. (1962). *The structure of scientific revolutions.* Chicago, IL: University of Chicago Press.

Lewin, K. (1935). *A dynamic theory of personality.* New York, NY: McGraw-Hill.

Schefler, I. (1972). Vision and revolution: A postscript on Kuhn. *Philosophy of Science, 39,* 366–374.

Smith, B. (1988). *Foundations of Gestalt theory.* Munich, Germany: Philosophia.

Verstegen, I. (2001). Gestalt psychologists on the Gestalt switch. *Il Cannocchiale Revista di studi filosofici, 2,* 3–9. Retrieved from http://sites.google.com/site/ianverstegen/gestaltswitch

Wittgenstein, L. (2001). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Oxford, England: Blackwell. (Original work published 1952)

Wright, E. (1992). Gestalt switching: Hanson, Aronson, and Harré. *Philosophy of Science, 59,* 480–486.

Given, Myth of the

The *myth of the given* is Wilfrid Sellars's name, introduced in his *Empiricism and the Philosophy of Mind,* for a broad epistemological strategy shared between a wide variety of empiricist traditions. The work was originally delivered as lectures at the University of London (Special Lectures on Philosophy for 1955–1956), on March 1, 8, and 15, 1956, under the title *The Myth of the Given: Three Lectures on Empiricism and the Philosophy of Mind.* Subsequent editions appeared.

The American philosopher Wilfrid Sellars (1912–1989) took as uncontroversial the insight that there is a distinction between seeing that something is the case and inferring from experience that it is the case. Empiricism in the grip of the myth of the given improperly transforms this fact into a foundationalist epistemology of immediacy, according to which there is some layer of entities of which we have direct receptive awareness—he names "sense contents, material objects, universals, propositions, real connections, first principles, even givenness itself" as contenders—that can function autonomously and directly as "data" from which we can infer. Although his primary targets were his recent predecessors—most noticeably sense-datum theorists such as the British philosophers G. E. Moore and Bertrand Russell—Sellars's critique of this myth of the given is, by his own description, analogous to Hegel's critique of immediacy (throughout Hegel's *Phenomenology of Spirit* but presumably most directly in the "Sense-Certainty" chapter of that work). Sellars makes it clear that he wants his critique to be broad in its historical target; at various points, he uses the myth of the given as a tool for criticizing not only the sense-datum theorists but classical empiricists of the 17th and 18th centuries, such as John Locke, Bishop Berkeley, and David Hume, and 20th-century logical positivists and behaviorists, such as Rudolf Carnap, A. J. Ayer, and Gilbert Ryle.

According to sense-datum theory, taken as a form of epistemological foundationalism, knowledge has two distinct sources: (1) sensory data, which are mind-dependent results of observation of which we are directly aware, and (2) a faculty of reason that draws conclusions from these empirical raw materials. In Section 6 of his work, Sellars argues that

sense-datum foundationalism is committed to the following "inconsistent triad":

A. *x* senses red sense content *s* entails *x* noninferentially knows that *s* is red.
B. The ability to sense sense contents is unacquired.
C. The ability to know facts of the form *x* is ø is acquired.

Without Proposition A, sensory data would not supply any premises for inference. Without B, sense-data would not be foundational. C asserts that the capacity to make conceptually articulated, truth-valuable judgments is something we develop rather than an innate capacity. Put together, the three propositions say that we have an unacquired capacity (to sense red sense contents) that entails that we are in a state that depends on our having an acquired capacity (to know the fact that *s* is red). Whether or not these count as technically inconsistent, they are certainly in strong tension with one another.

The sense-datum theorist needs A and B, but why must she be committed to C? A major project for Sellars (1997) in *Empiricism and the Philosophy of Mind* and elsewhere is to argue that the ability to make propositionally structured judgments is a holistically determined capacity that depends on our being embedded in a normatively structured "space of reasons" in the right way; it requires that we are capable of drawing inferential connections between those judgments and others and of making normative assessments of their justification:

> The essential point is that in characterizing an episode or a state as that of knowing, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says. (sec. 36)

We return to his reasons for thinking this below. But assuming for the moment that Sellars is right, such normative holism poses a serious challenge to any version of a foundational, autonomous given. If what is directly and noninferentially given in experience is not a conceptually articulated, truth-valuable content but rather a "mere look" or something of the sort, then it doesn't seem like the given can serve as an epistemic foundation justifying judgments. For such a thing doesn't have the right form to be a premise in an inference at all; its relationship to our conceptually articulated, propositionally formed beliefs seems to be merely causal.

Subsequent philosophers took similar positions. Donald Davidson embraces this conclusion, insisting that nothing can justify a belief except other beliefs, and hence in his view the (outside) world plays a merely causal role in constraining belief. In this case, the given is not functioning epistemically, as it is supposed to *ex hypothesi*. John McDowell argues that assigning observation a merely causal role leads to an unacceptable coherentist picture in which the whole is spinning in the void without friction from the world, as he famously puts it. Otherwise, a defender of a nonpropositional sensory given is stuck, claiming that a merely causal relationship between the world and our observational states can count as a normative justificatory relation, which seems to miss the holistic character of normativity. If, on the other hand, what is directly and noninferentially given in experience is something that is *already* propositionally formed, so that experience becomes inherently judgmental and we *see-as* in the first instance, then Sellars's holistic and normative account of judgment implies that it cannot function as relevantly "given" after all—we can't have an autonomous, innate capacity to make such judgments that doesn't depend on any other epistemic capacities. Either way, there is no given.

Sellars argues for both the holistic normative embeddedness of propositional judgment and the nonfoundational status of mere sensory contents in a variety of ways. The two arguments are combined most powerfully and memorably in his parable of John, the necktie salesman, who doesn't understand that colors look distorted in different lighting and can only be reliably detected under standard conditions. John thinks he sees that a necktie in his store is green, but his colleagues teach him that he needs to take the tie outside to see its blue color properly. In response to this training, he learns to say that the necktie only *looks* or *appears* green under artificial light but he can *see that* it is blue once he takes it outside. Sellars's point here is (at least) twofold. (1) If we think immediate sensory appearances are epistemically foundational, we might assume that *looks-talk* is prior to *seeing-that* talk. But in fact the logic of looks is *derivative upon* the logic of

seeing-that. It is only by *withholding* our judgment that we are justified in believing what we seem to see that the notion of mere looks or appearances gets a grip. (2) In turn, judgmental epistemic states such as seeings-that are in fact *normative* and *holistically determined* states. It is only once we can understand the conditions under which our judgments are justified and can draw a variety of inferential connections between these judgments—which requires knowing a variety of facts, such as what count as standard viewing conditions for colors—that we can properly count as *judging* that things are thus and so. Observational judgment does not *itself* require inference, but it depends on our already having reasonably sophisticated capacities to negotiate inferential and other connections in the space of reasons. Hence, propositionally structured seeings-that are not amenable to the kind of foundationalist analysis that those in search of an autonomous "given" seek. Sellars (1997) sums up thus:

> While [John's, the necktie salesman's, story] does not imply that one must have concepts before one has them, it does imply that one can have the concept of green only by having a whole battery of concepts of which it is one element. It implies that while the process of acquiring the concept green may—indeed does—involve a long history of acquiring piecemeal habits of response to various objects in various circumstances, there is an important sense in which one has no concept pertaining to the observable properties of physical objects in Space and Time unless one has them all—and, indeed, as we shall see, a great deal more besides. (sec. 19)

Sellars distinguishes two dimensions of epistemic dependence: inferential or evidentiary dependence, on the one hand, and what we might call a constitutive dependence, wherein the capacity to be in one epistemic state depends on the capacity to be in others. He takes the assumption that these track on another—that is, the assumption that one epistemic state presupposes another only if the second is its premise—to be a version of the myth of the given itself, as he explains in the following:

> One of the forms taken by the Myth of the Given is the idea that there is, indeed must be, a structure of particular matter of fact such that . . . each fact can not only be noninferentially known to be the case, but presupposes no other knowledge. . . . It might be thought that this is a redundancy, that knowledge . . . which logically presupposes knowledge of other facts must be inferential. This, however, as I hope to show, is itself an episode in the Myth. (sec. 8)

Observations have an asymmetric evidentiary relationship to indirect knowledge claims. I might claim to know various Galilean laws concerning the acceleration of balls on an inclined plane. Each of these generalizations are, along one dimension, based on a range of observations of the form "Ball *b* took *t* seconds to fall *n* inches down plane *p* with inclination *i*." But at the same time, my capacity for such observational knowledge is not something I have innately. For my sensory engagement with the experiment to count as *seeing that* the ball took *t* seconds to fall, I need to understand what it would mean for my observational state to be justified or not. Accordingly, I must understand that my observation is only reliable if the plane was reliably produced by skilled craftsmen, my clock is accurate, I am not intoxicated, there is adequate and nondistorting light in the room, and so on. Hence, I don't count as seeing *that* something is the case, in a way that is inferentially fecund, unless I know all these other facts as well. Thus, Sellars concludes that the ability to see that things are thus and so (or, more derivatively, that they look thus and so) is a capacity that requires already an ability to know that things are thus and so, even though along another logical dimension the former can sometimes count as evidence for the latter. While observational knowledge is noninferential, it is not *immediate* in the sense of being an autonomous source of information that we could have prior to and independently of any acquired epistemic capacities.

Sellars claims to secure an epistemological position liberated from the myth of the given, in which each of the following is true:

1. We know many things about the empirical world.
2. Our knowledge is constrained by our noninferential observations of that world.
3. This constraint is normative: To see that *x* is red is to *know* that *x* is red and to be in a position to justify my belief that *x* is red; it does not merely cause me to have that belief.
4. Observation is at one and the same time a noninferential source of knowledge and not a source of an autonomous given foundation for knowledge.

The question of how to develop the details of a post-Sellarsian epistemology that avoids the myth of the given has been central for philosophers such as Robert Brandom, John McDowell, and Donald Davidson. In turn, some contemporary philosophers, such as Richard Heck and Michael Luntley, have challenged Sellars and his descendants by accusing them of being committed to an overly conceptual, propositional picture of what content has to be like for it to be epistemically fecund; they have defended instead an epistemic role for nonconceptual content.

Rebecca Kukla and Mark Lance

See also Empiricism; Epistemology; Hegelianism and Contemporary Epistemology; Holism, in the Philosophy of Language; Inferentialism; Nonconceptual Content; Observation and Theory-Ladenness; Truth, Philosophical Theories of

Further Readings

Alston, W. P. (2002). Sellars and the "myth of the given." *Philosophy and Phenomenological Research, 65,* 69–86.

Brandom, R. (2002). *Tales of the mighty dead: Historical essays in the metaphysics of intentionality.* Cambridge, MA: Harvard University Press.

Chalmers, D. J. (2003). The content and epistemology of phenomenal belief. In Q. Smith & A. Jokic (Eds.), *Consciousness: New philosophical perspectives* (pp. 220–272). Oxford, England: Oxford University Press.

Chisholm, R. (2000). The myth of the given. In E. Sosa & J. Kim (Eds.), *Epistemology: An anthology* (pp. 107–120). Oxford, England: Blackwell.

Davidson, D. (1986). A coherence theory of truth and knowledge. In E. LePore (Ed.), *Truth and interpretation: Perspectives on the philosophy of Donald Davidson* (pp. 307–319). Oxford, England: Blackwell.

DeVries, W., & Triplett, T. (2000). *Knowledge, mind and the given: Reading Wilfried Sellars's "Empiricism and the Philosophy of Mind."* Indianapolis, IN: Hackett.

Echelbarger, C. (1974). Sellars on thinking and the myth of the given. *Philosophical Studies, 25,* 231–246.

Heck, R. (2007). Are there different kinds of content? In J. Cohen & B. McLaughlin (Eds.), *Contemporary debates in the philosophy of mind* (pp. 117–138). Oxford, England: Blackwell.

Kukla, R. (2000). Myth, memory, and misrecognition in Sellars' "Empiricism and the Philosophy of Mind." *Philosophical Studies, 101,* 161–211.

Luntley, M. (2003). Nonconceptual content and the sound of music. *Mind & Language, 18,* 402–426.

McDowell, J. (1994). *Mind and world.* Cambridge, MA: Harvard University Press.

McDowell, J. (1998). Having the world in view: Sellars, Kant, and intentionality. *Journal of Philosophy, 95,* 431–491.

O'Shea, J. (2002). Revisiting Sellars on the myth of the given. *International Journal of Philosophical Studies, 10,* 490–503.

Robinson, W. S. (1975). The legend of the given. In H.-N. Castañeda (Ed.), *Action, knowledge and reality* (pp. 83–108). Indianapolis, IN: Bobbs-Merrill.

Rosenberg, J. F. (2007). *Wilfrid Sellars: Fusing the images.* Oxford, England: Oxford University Press.

Sellars, W. (1997). *Empiricism and the philosophy of mind* (with an introduction by R. Rorty and a study guide by R. Brandom). Cambridge, MA: Harvard University Press.

Sosa, E. (1997). Mythology of the given. *History of Philosophy Quarterly, 14,* 275–286.

Goal-Directedness

Goal-directedness is not a fully systematized topic. It includes several subdomains, with open issues, some contradictory definitions, different approaches, and various research trends. This entry mainly focuses on foundational issues, leaving aside some important and growing domains, such as studies and models about the "control" of intentional motor action, experimental evidence, and neural correlates (e.g., Marc Jeannerod's work); the crucial research on *intention recognition, mind reading, simulation, mirror neurons,* and so on (as, e.g., in the work of Giacomo Rizzolatti or Vittorio Gallese); or, more generally, the theory of *intentional action.* All these topics are addressed in other entries in this encyclopedia.

The study of goal-directedness is crucial for social action, since the latter must unavoidably be construed in terms of its underlying ends and functions.

The Return of the "Final Causes"

The year 1943 marked a turning point in science, with the simultaneous publication of the works of Kenneth Craik, Warren McCulloch, Walter Pitts, Arturo Rosenblueth, Norbert Wiener, and Julian Bigelow. Science resumed the concept of *end*—that is, of final cause or teleology—which had been driven out of it and still continues to elicit misunderstandings and problems.

"Teleological" Versus "Teleonomic" Systems

There are several uses of the term *goal directed* (GD). The most important are (a) behavior with a cybernetic goal-based control system and (b) purposive/functional behavior not internally regulated by a goal representation. So "goal directed" is used as a synonym for "finalistic" or "teleological." Analogously, "goal" is used to denote either (a) an internal guiding anticipatory representation or the functional/selective effect of the action or system feature responsible for its reproduction and maintenance (it must be remarked that these two notions are frequently mixed up in the relevant literature) or (b) the external object or "target" of an action.

It is better to use the biologist Ernst Mayr's terminology, which clearly disentangles two kinds of GD systems, distinguishing "teleonomy" from general "teleology." This distinction between two kinds of finalistic behavior is really crucial: (1) the "goal-governed" (GG) or "goal-driven" systems (controlled from inside by an internal representation of the goal state or set point) and (2) the merely "goal-oriented" (GO) ones, where the behavior has a specific goal or, better, a function to be realized, which, however, is not internally represented and anticipated and guiding/governing the performance. These two kinds of finalistic behavior are not one and the same things; in evolutionary terms, the goal-based internal control of behavior may be viewed as a specific—evolutionarily advanced—way of providing teleology to behavior; it is the internal flexible implementation of its adaptive function.

Internally Represented Goals Versus Functional Results

There is also a distinction between internally represented "goals" and goals as functional results. Still, these two are frequently mixed up. For example, after correctly defining *goal* as a cybernetic representation, some scholars ascribe the goal of "survival" to living systems. This clearly is a high-level "function" of living systems, not a represented and "pursued" goal. Similarly, several authors of emotion theory, while correctly claiming that there are emotions only because there are goals (since emotions are signals about the possible realization or frustration of relevant goals), do not carefully distinguish between emotions "about" internally represented and pursued goals (e.g., joy, envy, guilt, etc.) and emotions that are simpler and merely reactive affective states whose goal is not represented.

This distinction within goal-directedness between GG and GO behavior also applies to social action. "Cooperation" among insects is not in view of some represented goal. Humans can cooperate for realizing a mentally shared goal and formulate the goal of jointly acting, or a common plan. The same holds for "communication," which necessarily is teleological. In many animals communicating is not a case of goal-directedness, while in humans it usually is intentional, but linguistic communication is not necessarily intentional.

Goal-Directedness and Anticipation

The distinction between two kinds of "anticipation" and "anticipatory" behaviors/systems is crucial in goal-directedness studies. We can distinguish between

- *cognitive anticipation,* where the system activates/builds the "expectation," the representation of the future event, and behaves accordingly, and
- *functional* (but not representational) *anticipation,* where the behavior just has the function of (has been selected or learned for) coping with an impending event and is elicited by some announcing signal.

In this respect, GG behavior exploits a cognitive anticipatory device, a representation of the future, of *what doesn't exist* (yet) or is *not perceived*.

However, not all cognitive-anticipatory systems/behaviors are GG. There are simpler, and probably more primitive, mechanisms exploiting anticipatory representations ("expectations" proper), which do not imply goals in strict control theory terms. They are the "anticipatory classifiers": a simple rule (Condition ⇒ Action) plus the prediction of the relevant outcome (expectation): C ⇒ Act + Exp. The animal *reacts* to the stimuli, performs the action, and is expecting the reward and checking for it. The confirmation of that prediction will feed back on the strength, activation probability, and persistence of the rule.

Some kinds of behavior (e.g., a behavior resulting from "instrumental learning") may look like a real GG act, in that it is "finalistic" and there is an anticipated mental representation of the outcome playing

a role in that behavior. However, GG behavior is not necessarily its proximate mechanism in this case; that behavior might also be explained as anticipatory-classifier based. The internal stimulus (e.g., a drive like hunger) and the external one (a particular environment, e.g., a red button—i.e., specified "conditions") activate the right "rule" (classifiers), that is, the associated action/response, and also the expectation of the outcome (food). In GG behavior, what activates the behavior is not the "condition" (a stimulus) but the mismatch between the goal representation and the perceptual/doxastic (belief centered) representation of the current state of the world.

The TOTE Model

This is the psychological version of the cybernetic model of goal-directedness—the TOTE (Test–Operation–Test–Exit) model of Miller, Galanter, and Pribram.

In a sense, the system starts from the goal representation, which is not yet or necessarily an "expectation" about what will happen (it might already be true, or it might be assumed as impossible, i.e., never true), and uses it (a) for evaluating the current state of the world, (b) for activating or even searching in memory for the appropriate action (an action with such an "expected" result), (c) for planning by finding context-related subgoals, (d) for monitoring the execution and adjusting it, (e) for terminating the action (when the outcome matches the goal), and (f) in affective systems, in particular, for eliciting the appropriate felt or affective response (pleasant/unpleasant depending on the "success" or "frustration") or emotion, which are always goal related and usually caused by the realization or frustration of true internal goals.

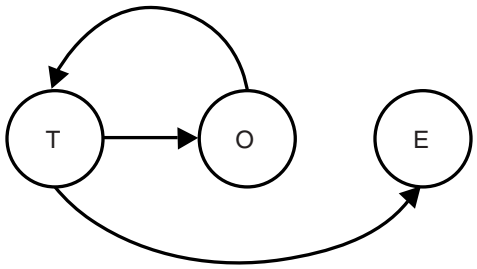


Figure 1 TOTE Model (Test–Operation–Test–Exit)

Source: Cristiano Castelfranchi.

Problems With the Cybernetic Definition

Some problems with the original cybernetic definition of purposive (GD) behavior are as follows:

1. A basic problem was due to the strictly behavioristic approach of Norbert Wiener and associates, aimed at avoiding the necessary "internal" (mental) representation, hence Taylor's objection that this mechanism cannot capture real (human) purposive behavior, based on "desires" or "beliefs," that is, on "subjective" states. Also, the ideas of "final cause," of the backward causation, loop, circular causation, and so on, were eliciting a lot of criticism about the violation of natural causality or about its interpretation.
2. The cybernetic model was too primitive: There was no subgoaling but just "Mismatch ⇒ Action activation/execution"; there was no action selection, just a unique goal; there was no managing of several simultaneously active goals (possibly incompatible with each other), and thus a model of goal selection or even true decision/deliberation was lacking; similarly, there was also no model of goal dynamics, activation, abandon, and so on, or of goal hierarchy (see the next problem).
3. There was no concern for the "value" or hierarchy of more than one goal (i.e., questions related to the level of activation, felt intensity, calculated utility, etc.), something unavoidable when one has several goals and has to "choose" among them.
4. In the original cybernetic-inspired goal models imported into psychology, goals were only conceived as represented states of the external world, whereas we know that specific cybernetic representations (motor or proprioceptive) and feedbacks control the execution of the action from inside. Goals are not only about the "external" (final) result on the world, but they are about the internal effects of the movements, step by step, and they finally control and adjust the behavior accordingly.
5. Goals useful for psychology (and artificial intelligence) are seen as explicit (and manipulable) representations that can be generatively produced, changed, analyzed, combined, and compared; one can "reason"

about them and manipulate them. They can be either sensory-motor or symbolic representations (e.g., propositional attitudes), but in all cases they are understood as real mental objects to be managed. A problem, however, is this: Not all the "cybernetic" goals are "mental," control our "action," or assume the subjective forms of motor control, of desire, intention, or objective, or of a project to be pursued, and so forth. There are also real cybernetic (set-point based) homeostatic regulations in our body (e.g., maintaining our body temperature around 35–40°C) that are, however, not mentally represented; that is, they are not "subjective" goals and do not strictly speaking regulate our "actions," when the latter are taken as appropriate to intentional agency or conscious social acting. It might also be the case that some function (e.g., "survival") is implemented not just in stimulus–response rules but in internal set-point goals (e.g., "be far from the threat"); however, this is not a representation of the end "survival."

Properties and Functions of Goals

Motives

In GG behavior, taken as "motivated" behavior, the following basic idea is maintained: A mental goal "motivates" the action. Spelled out, this means that

- the action wouldn't be performed if not in view of that result—that is, the goal is *necessary* for choosing and performing that action;
- the behavior is stopped if there is no such expectation, either because the expected result is already realized or because it is considered impossible;
- the goal bestows "value" to the action, and the action (in competition with other possible actions) receives "value" from its goals—we choose between goals (outcomes), not between actions; and
- the goal is the explanation, being the "reason" of a given action—that is, its "motive."

Real Goals

The second issue to be clear about is specifying people's real goals. The explicit identification of the subject's goals is thus fundamental for understanding GD (and especially GG) behavior. But this contrasts with certain developments in the social sciences, in particular with the divorce of classical economics from psychology (at least since Wilfredo Pareto). One ought to be able to explain why people go to vote or have the so-called sunk-costs bias; taking into account the actors' subjective goals, the resulting behavior may be seen as antieconomic. Yet it cannot be seen as "irrational" relative to the given goals. Normative economics cannot prescribe to people the "right" goals to have.

Hierarchical Organization

Goals have a "hierarchical" organization. Over any action, there is a means–end goal chain: A goal has subgoals and superordinate goals.

For realizing a given goal, we need some action; performing it becomes the subgoal. But an action requires specific skills and external conditions: They generate another subgoal level, and so on. "Planning" therefore is necessary: To put Cube A on Cube B, B must be "free" (i.e., this is a condition); if it is not free, we have to act to make it free (subgoal, subaction). The "deep structure" of action is not therefore a mere "sequence." Actions are assembled by their goals: Actions cooperate in "plans"; actions are composed of other, simpler and subordinate actions. We are both able to build new plans for solving new problems and also memorize and reuse complex actions and plan structures, both in individual and in collective actions—that is, create and use scripts. Goal-based scripts and plans (memorized and shared, or created) are not just behavioral sequences: They "coordinate" social action.

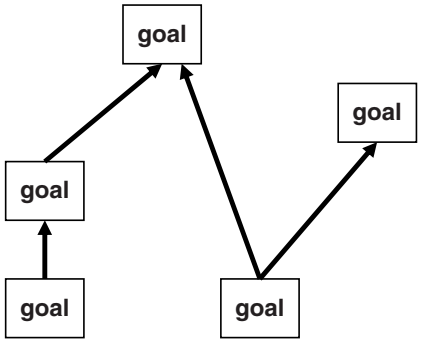


Figure 2 Goal Hierarchy

Source: Author.

Abstraction

Current neural and experimental studies are rightly focused (also thanks to the current impact of the "embodied" view of mind) on motor action control and on "concrete" goals (sensory-motor representations). However, we must not forget that "action" is not equivalent to "motor action" and "goals" are not equal to "perceptual goals." An action can be "abstract"; it can just be *to bring it about that . . .* or *to see to it that . . .*, without specifying "how"—that is, without specifying its motor implementation, the subactions that will be its vehicle. For example, to *offend* is an action, but did we write or say something offensive to that guy, show him the finger, or simply intentionally ignore him? Similarly, a goal can be abstract, like *to make money* or *to win some competition*. Abstract goals raise an important problem: How can it be checked whether they have been achieved or not? On the basis of which particular set of specified evidence and beliefs? Abstract goals therefore have to be "translated" into more concrete cues and signals of their realization.

"Goals" or "Desires and Intentions"?

An important tradition in philosophy, psychology, artificial intelligence, and logic is the so-called BDI (beliefs–desires–intentions) approach, where the behavior ("action") is clearly GG and purposive (or, rather, intentional). However, scholars in that tradition do not use a unifying category of "goal" but introduce two motivational independent "primitives": *desires* (not necessarily pursued, possibly contradictory, etc.), on the one hand, and *intentions* (chosen/planned, coherent, implying the agent's commitment, etc.), on the other. They add important distinctions like "intention that" versus "intention to do," "intention in action" versus "prior intention" (as in John Searle's terminology), and "future-directed intention" versus "commitment."

Is Internal Goal-Directedness Equal to "Selfishness"?

There is frequently a dangerous confusion between goal-driven or self-motivated agents and "selfish" ones. Self-interested or self-motivated agents are simply *endowed with and guided by their own internal goals*. The fact that they are "autonomous," necessarily driven by their internal ends and taking into account the value of their own motives, does *not* make them "selfish." This is a matter of motives, not of mechanisms. They may have any kind of motive: pecuniary or in general economic, moral, esthetic, prosocial, altruistic, self-sacrificial, and so on. Whether they are selfish or not depends just on the basis of the *specific* motives they have and prefer, not on the basis of their being driven by internal or endogenous motives.

Is Behavior Goal Directed/Goal Governed or Pleasure/Utility Directed?

Do we have several final, independent goals (motivations) or just one dominating and monarchic, as it were, goal, like "utility maximization" or "pleasure"? Maximizing utility is for sure a *function* of the goal-processing mechanisms, but is it a real represented "goal" we always pursue, while our specific motives would just be instrumental subgoals? The same goes for pleasure. Is it our overarching (unique) goal? Is it really and always our explicit goal? Or is it just a fundamental feedback of purposive behavior, a reward for learning goals? It is simply false that any goal achievement implies a pleasure experience; what about subgoals like putting a stopper on a bottle or pushing the button of the elevator?

A Paradox of Goal-Directedness: Akrasia

How is it possible that we "choose" and pursue some goal that we would not like to pursue, that we do not consider to be in our best interest, or that we self-consciously consider to be immoral and that it ought not to be pursued? (This is the ancient Greek philosophical notion of *akrasia,* which Socrates, Plato, and Aristotle struggled with, giving distinct answers.) In modern terms, the question may be phrased as follows. Is there a real choice, a real conflict between two intentions (one perhaps unconscious)? Is there an unconscious "rational" choice of our best preference, with calculation of secondary advantages? Or is it a fight between our intentional control system and other, more impulsive, automatic systems, which, though unintentional, may control our purposive behavior? (See also the "dual-processing" literature.)

Conclusion: "Goal-Directedness" as a Frame

It can safely be maintained that not only is "behavior" or "action" GD and GG (and social behavior and relations too) but also cognition itself. Our

beliefs (and their acquisition, elaboration, and use) are in relation to and a function of our mental goals (occurrent or potential). Our reasoning is an instrument for goal selection, planning, and achievement. Learning, too, is driven by "success" or "failure." Also, emotions are goal guardians and signifying signals of the achievement or thwarting of our goals, as well as activators of goals. Goal-directedness and management of goals is thus the real structure of behavior and mind. Mind and action must be principally read by means of a theory of their goal-directedness. Moreover, to understand and model social attitudes, social actions, and social structures, we have to model their underlying goals, intentions, and functions.

Cristiano Castelfranchi

See also Action, Philosophical Theory of; Agency; Artificial Intelligence; Cognitive Sciences; Complexity; Complexity and the Social Sciences; Cooperation/Coordination; Emergence; Feedback Mechanisms and Self-Regulatory Processes in the Social Sciences; Intention, Social Psychology of; Intentionality; Laws Versus Teleology; Systems Theory; Unconscious Social Behavior

Further Readings

Aarts, H., & Elliot, A. (Eds.). (2011). *Goal-directed behavior.* London, England: Taylor & Francis.

Bratman, M. E. (1999). *Intention, plans, and practical reason.* Stanford, CA: CSLI. (Original work published 1987)

Castelfranchi, C. (2001). The theory of social functions. *Journal of Cognitive Systems Research, 2,* 5–38.

Cordeschi, R. (2002). *The discovery of the artificial: Behaviour, mind and machines before and beyond cybernetics.* Dordrecht, Netherlands: Kluwer Academic.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York, NY: Holt, Rinehart & Winston.

Moskowitz, G. B., & Grant, H. (Eds.). (2009). *The psychology of goals.* New York, NY: Guilford Press.

Okrent, M. (2007). *Rational animals: The teleological roots of intentionality.* Athens: Ohio University Press.

Paglieri, F., & Castelfranchi, C. (2008). More than control freaks: Evaluative and motivational functions of goals. *Behavioral and Brain Sciences, 31*(1), 35–36.

Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, purpose and teleology. *Philosophy of Science, 10*(1), 18–24.

Governmentality and Regime

Governmentality is a term, most often derived from Michel Foucault's lectures, that focuses on the field of practices, techniques, and forms of knowledge through which we govern in modern societies. It concerns, but is not restricted to, the political government of the state. It has several meanings and usages. First, governmentality refers to the chronological process by which the state comes to accept responsibility for governing the population, economy, and civil society and within which economic knowledge will assume an important role. Second, it is a field of power relations that is concerned with governing through the self-government of individual and collective actors. Third, governmentality is a form of analysis that links the study of the government of self and the government of others and is a dimension of Foucault's broader project of making intelligible the historical conditions of contemporary experience. Finally, governmentality studies have become a distinctive subfield of study in the social sciences.

Regime is a term meaning the more or less organized ways we produce truth and knowledge (as in "regime of truth") and seek to govern others and ourselves ("regime of practices" and "regime of government").

This entry examines the meaning of "governmentality" in the work of Foucault, the impact of the term in governmentality studies, and key issues pertaining to the concept of regime.

Foucault and Governmentality

The central sources for Foucault's conception of governmentality are his 1978 and 1979 published lectures (*Security, Territory, Population* and *The Birth of Biopolitics*), in which he introduces the concept as an object, a chronological process, and a field. As an object, it is a specific historical "ensemble" that takes population as its target, political economy as its form of knowledge, and the apparatus (or *dispositif*) of security as its technical means. Chronologically, it is the process of how the "state of justice" and the "administrative state" in Europe become "governmentalized" beginning in the 18th century. As a field, it is a kind of *power,* distinct from others such as sovereignty and discipline, that has become preeminent. This kind of power is closely related to the emergence of what Foucault calls a "liberal art of

government," which emerges from the critique of the claims of state sovereignty to omniscience and omnicompetence in early-modern doctrines of reason of state and police. More recently, the welfare state has been transformed partially as a result of the "neoliberal" suspicion of the pretensions of state intervention, state regulation, economic planning, and social provision. Liberalism can be characterized as anxiety that one always governs too much.

In the same lectures, Foucault traces the emergence of this modern governmentality from the Christian pastorate and stresses the idea of the government of conduct. Conduct here is not only the art or activity of conducting but also the way in which one conducts oneself. Government, in a very broad sense, becomes the "conduct of conduct." It is possible to talk of governing families, souls, populations, states, and even oneself.

In later lectures in 1982 and 1983 (*The Hermeneutics of the Subject* and *The Government of Self and Others*), Foucault is concerned less with modern power and more with ancient ethics and ascetics. Here, he expands upon a relational conception of governmentality, which is capable of linking and analyzing power relations, governing the self and others, the relation to oneself, as a chain that makes it possible to connect political government with ethical self-government, politics to ethics. A further expansion of the concept occurs when Foucault places the analysis of governmentality in relation to other dimensions of his work. His general approach is to uncover the conditions and practices through which apparently given, stable, and universal entities emerge. Here, "procedures of governmentality" move beyond the history of domination and state in the same way as "regimes of veridiction" displace the history of knowledge and a "pragmatics of the self" replaces the history of subjectivity.

Regime

Foucault applied the term *regime* throughout his work to refer to more or less organized ways of producing truth and knowledge or of governing individual and collective actors for particular ends. At its most abstract, he discusses "regimes of practices" as organized ways of governing conduct, on the one hand, and sites for the production of knowledge, on the other. In *Questions of Method,* for instance, he argues that his principal concerns are the regularities, logic, and strategy embedded in a regime of practice, these practices being shaped through programs and rationalities concerned with conduct and giving rise to certain kinds of knowledge. A regime of practices is a way then to characterize how humans seek security, punish, cure or care, and so forth.

However, there is another term, *dispositif,* which has given rise to a broad discussion and which has been translated as apparatus, assemblage, device, and dispositive. Foucault uses this term to discuss regimes of law and sovereignty, discipline, sexuality, and security. Other contemporary philosophers such as Gilles Deleuze and Giorgio Agamben have sought to install this notion of dispositive as a kind of master key to Foucault's work. Like the concept of regime, it stresses the articulation between rationalities, technologies, fields of visibility, forms of subjectivity, and ends sought, which appear in a given domain of practice. This term delineates a greater immutability and breadth of impact than the idea of regime.

Governmentality Studies

In Anglophonic countries and more recently Germany and Scandinavia, Foucault's writings on governmentality have given rise to a literature that is capable of studying anything from social welfare practices to human resource management, corporate social responsibility to risk, and criminality. This has been called *governmentality studies,* although most of its practitioners would have reservations about the term. Such studies have often concerned themselves with "advanced liberalism," less as an expression of neoliberal philosophies and more as a series of techniques, rationalities, and devices that bear a mutual resemblance. Questions of the dominant logics of government of and by the state remain matters of debate and investigation, particularly the relationship of governmentality to the sovereign powers of the state and the international state system.

Mitchell Dean

See also Foucault's Thought; Genealogy; Power; Structuralism and Poststructuralism

Further Readings

Burchell, G., Gordon, C., & Miller, P. (Eds.). (1991). *The Foucault effect: Studies in governmentality.* London, England: Harvester/Wheatsheaf.

Bussolini, J. (2010). What is a dispositive? *Foucault Studies, 10,* 85–107.

Dean, M. (2010). *Governmentality: Power and rule in modern* society (2nd rev. ed.). Thousand Oaks, CA: Sage.
Foucault, M. (1991). *The Foucault effect: Studies in governmentality* (Graham Burchell & Colin Gordon, Eds.). Chicago, IL: University of Chicago Press.
Foucault, M. (2006). *The hermeneutics of the subject: Lectures at the Collège de France 1981–1982* (Graham Burchell, Trans.; Frederic Gros, Ed.). London, England: Picador.
Rose, N. (1999). *Powers of freedom: Reframing political thought*. Cambridge, England: Cambridge University Press.
Sennellart, M. (2007). Course context. In M. Foucault (Ed.), *Security, territory, population: Lectures at the Collège de France 1977–1978* (pp. 369–391). London, England: Palgrave.

Grounded Cognition and Social Interaction

This entry explains novel approaches of grounded cognition and mental representation where sensorimotor simulation is central, as well as how these relate to accounts of *social interaction* in which embodied cognition of higher social concepts or the grounded nature of affect and emotion are vital themes. An introduction to grounded cognition is followed by a discussion of social interaction informed by grounded cognition approaches.

Grounded Cognition

Cognition is the ability to mentally represent the world and to manipulate those representations in order to understand, remember, use language, plan actions, and reason.

Sensorimotor Simulation

Whereas more traditional theories in cognitive science assumed that such mental representations consisted of abstract symbols, the psychologist Lawrence Barsalou, building on prior philosophical ideas, proposed that representations consist of sensorimotor simulations. On this account, the *perceptual symbols theory,* cognition shares processing mechanisms and neural systems with perception and action.

During an experience, neural patterns are activated in modality-specific areas for sensory and motor processing. Higher-level association areas (*convergence zones* in Antonio Damasio's words) capture these activation patterns at different levels. Association areas in modality-specific areas capture activation patterns within modalities, and higher-level areas integrate patterns from different modalities.

Representation is essentially the reenactment of previous experiences. Higher-level association patterns activate lower-level association areas, which activate patterns in sensorimotor areas. Rather than an exact replica of experiences, however, these patterns are distorted and represent only partial experiences. In Barsalou's theory, simulators capture patterns of activation for a particular category of experiences. As a result, they represent a distributed pattern of experiences with a concept such as *chair* or *apple*. Simulation of experiences is dynamic and flexible and can even represent imaginary events.

Research has shown that representations are organized along sensorimotor modalities and contain modality-specific information. Representation and perception share processing mechanisms and have been shown to support or interfere with each other. For example, (a) after a shift in sensory modality, similar costs in processing occur between and within perceptual processing and mental representations, and (b) people recognize object pictures faster when their orientation, shape, or motion matches the values implied by language. Neuroscientific studies have indicated that overlapping brain areas are involved in representation and sensorimotor processing.

Cognition for Action

A related but slightly different view, proposed by Arthur Glenberg, is that cognition is for action. On this account, the function of memory is to support actions. This is achieved by *meshing*, a process that integrates properties of the environment, embodied memories, and properties of the body into potential action patterns. The meshing process serves both to understand the current environment and to mentally represent situations.

Social Interaction

Research on social interaction has long assumed that the body, and its mental representation, is closely tied to the processing of social and emotional information. In fact, the central concept in social psychology—attitude—has been thought more than a century ago as grounded in motor dispositions. Thus, Charles Darwin proposed that attitudes are collections of motor behaviors,

and Francis Galton thought that attitudes consist of *body* inclinations. Research on the role of the body flourished since then, with researchers showing, for example, that (a) nodding the head, as in agreement, while listening to persuasive messages leads to more positive attitudes toward the message than shaking the head, as in disagreement, and (b) objects associated with approach actions (e.g., arm flexion) are subsequently evaluated better than objects associated with avoidance actions (e.g., arm extension). Such phenomena fit the grounded cognition framework, which assumes that bottom-up bodily states are an intrinsic component of social information processing.

Embodiment of Higher-Order Social Concepts

Crucially, the grounded cognition framework goes beyond simple concepts and holds that high-order mental processes operate on perceptual symbols. Thus, even complex concepts involve partial reactivations of the sensorimotor states that occur during an individual's original experience with the world, as well as simulation of new sensorimotor experiences. Evidence that such processing occurs for complex social concepts has been obtained in several domains. For example, (a) making a fist influences people's processing of words related to the concept of power and (b) washing hands changes people's perception of moral responsibility or luck.

Simulating Affect/Emotion

The grounded cognition approach naturally applies in the domain of affect and emotion, where it is straightforward to conceive of emotion knowledge as a process of internal simulation of an emotional experience. Clearly, these simulations may not constitute full-blown emotions, must not simulate every aspect of emotion, and may not even be conscious. Nevertheless, such simulations can generate enough sensorimotor information to support conceptual processing. Evidence for simulation of emotion knowledge ranges from (a) activity of similar brain networks when people experience actual disgust and when people simply think about disgust, even in its abstract moral form, to (b) facial muscular activity when people abstractly think about concepts such as anger and happiness, to (c) activation of general introspective processes after people have been focused to think about abstract emotional concepts from an experiential perspective. Again, these mind-to-body findings are consistent with body-to-mind evidence that manipulating the motor processes can change the experience of the emotion, as well as its perception. Thus, people in a slumped posture feel less proud than people in an upright posture, and blocking expressions of happiness interferes with recognition of that emotion.

Imitation

A classic finding is that people imitate emotional expressions, gestures, and mannerisms even when not encouraged to do so. This phenomenon of spontaneous imitation can be understood by theories of grounded cognition, which view it as part of the bodily reenactment of the experience of the other's state. In fact, there is evidence that the lack of imitation may cause problems with social interaction.

Conclusion

The grounded cognition framework, with its proposal that mental processing is simulated experience, provides an essential theoretical account for understanding the mind. On this account, social interactions are supported by simulations of emotions, motor actions, and other people's states. Thus, social behavior is grounded in the body.

Diane Pecher and Piotr Winkielman

See also Concepts; Embodied Cognition; Emotions; Joint Attention and Social Cognition; Situated Cognition; Social Cognition; Social Neuroscience; Social Perception

Further Readings

Barsalou, L. W. (1999). Perceptual symbol system. *Behavioral and Brain Sciences, 22,* 577–660.

Damasio, A. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation Spring, 1*(1), 123–132.

Glenberg, A. M., Jaworski, B., Rischal, M., & Levin, J. R. (2007). What brains are for: Action, meaning, and reading comprehension. In D. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 221–240). Mahwah, NJ: Lawrence Erlbaum.

Niedenthal, P. M., Barsalou, L., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9,* 184–211.

Pecher, D., & Zwaan, R. A. (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking.* Cambridge, England: Cambridge University Press.

Group Beliefs

A group belief is a belief held by a group of two or more people, such as a couple, a club, a corporation, or a state. The question of whether groups may literally be said to have beliefs, and if so when, has significant implications for explanation and evaluation of group and institutional behavior in the social sciences and in ethics. If groups have beliefs, then their behavior may be explained and predicted by knowledge of what those beliefs are. When group beliefs lead the group to harm others or the environment, it may be appropriate to hold the group as a whole morally responsible.

This entry focuses primarily on the *ontological* debate about the existence and nature of group beliefs. First, the early division between holists and methodological individualists is briefly discussed; then, the current debates within the field of philosophy about the proper analysis of group beliefs are surveyed. The entry concludes with a brief discussion of recent work on group beliefs within epistemology.

Consider the following examples of group beliefs:

> We believe that Susan is the best candidate for the position.
>
> British Petroleum believes that technology is key in tackling the challenges faced by the biofuel industry.
>
> Mathematicians believe that Fermat's Last Theorem has been proved.

We often attribute beliefs to groups in order to explain the actions of individual members and the group as a whole. There is disagreement, however, about whether or not groups literally have beliefs. Some argue that such statements of group belief are merely a figurative way of speaking that should not be taken literally. Even among those who hold that there are genuine group beliefs, there is disagreement about what conditions must be satisfied for a group to have a belief.

Group Minds and Methodological Individualism

The attribution of beliefs to groups as a way to explain individual and group behavior was prevalent in sociology and psychology at the turn of the 19th century. Émile Durkheim, for instance, held that we could explain individual behavior by looking at the "collective consciousness" of the society of which the person is a part. Later theorists rejected this "holist" view as implying the existence of a metaphysically mysterious "group mind." According to *methodological individualism,* which arose early in the 20th century, we should explain all social phenomena in terms of the beliefs and behavior of individuals. On this "summative" view of group beliefs, statements of group belief are equivalent to statements about the beliefs of the *sum* of the members of the group.

Rejectionists and Believers

It had to wait until the late 20th century for a serious program of analyzing group concepts to begin. Such analyses are the focus of philosophers working on "collective intentionality," most notably Raimo Tuomela and Margaret Gilbert. Theorists of collective intentionality seek to analyze and understand how groups of individuals can share intentional states such as intentions to act, beliefs, and, in some cases, even emotions or perceptions. Tuomela and Gilbert each developed a theory of collective intentionality that rejects the assumptions of the methodological individualists, arguing that collective intentions are not simply a "sum" of individual intentions. In line with this, they each proposed a nonsummative account of group belief.

In response to such so-called believers, who claim that group beliefs exist, "rejectionists" have argued that there are no group beliefs. Rejectionists claim that groups may *accept* propositions but they cannot believe them. Rejectionists point out that there are a number of differences between beliefs and acceptances, and they argue that all group "beliefs" have the features of acceptances and frequently fail to have the features of genuine beliefs.

Table 1 Beliefs and Acceptances Compared

Beliefs	*Acceptances*
Involuntary	Voluntary
Aim at truth	Aim at pragmatic success
Follow evidence	Follow interests and desires
Ideally coherent	Allow for contradiction
Come in degrees	Categorical

Source: Kay Mathiesen

Interpretation or Agreement

Among "believers" who accept the idea that there are genuine group beliefs, different accounts are offered of the existence conditions for group belief. On the "interpretational" approach, advocated by Deborah Tollefsen, a group has a belief when the activities of the group and its members can be fruitfully explained by attributing beliefs to them. This approach is inspired by the "intentional stance" theory of intentional states, which holds that an entity has a belief just if we can give an illuminating explanation of its behavior by attributing a belief to it.

On the "agreement" approach, for there to be a group belief, the members of the group must each accept the proposition that is believed. Most notably, Margaret Gilbert's "plural subject" account proposes that a group forms a belief just when the members of the group "jointly accept" the proposition that expresses that belief. Once the members have jointly accepted a proposition, they are committed "as a body" to that proposition. This means that each individual member of the group is obligated to act and speak consistently with the group belief when they are acting as members of the group.

Extended Minds and Collective Epistemology

Recent work on group belief has extended beyond the original debate in social ontology to philosophy of mind and epistemology. The theory of the extended mind, developed by Andy Clark, brings a new perspective to the question of group beliefs. According to the extended mind hypothesis, minds are not contained within single individuals but can be spread across two or more persons who share cognitive resources. The emerging field of collective epistemology asks different questions about group belief than the ontologist, such as "Can groups have beliefs that may form the basis of knowledge?" If so, are groups knowers in their own right, of equal importance with individual knowers?

Kay Mathiesen

See also Collective Agents; Collective Intentionality; Distributed Cognition and Extended-Mind Theory; Group Mind; Holism, in the Social Sciences; Individualism, Methodological; Joint Attention and Social Cognition; Mutual Beliefs; Plural Subjects; Social Ontology, Recent Theories of; We-Mode, Tuomela's Theory of

Further Readings

Gilbert, M. (1987). Modeling collective belief. In *Living together: Rationality, sociality, and obligation* (pp. 195–214). Lanham, MD: Rowman & Littlefield.

Schmid, H. B., Sirtes, D., & Weber, M. (2011). *Collective epistemology.* New Brunswick, NJ: Ontos.

Tuomela, R. (1992). Group beliefs. *Synthese, 91,* 285–318.

Tuomela, R., Preyer, G., & Peter. G. (Eds.). (2003). *Understanding the Social II: Philosophy of sociality* [Special issue]. *Protosociology, 18–19* (Part IV).

Group Identity

In the social sciences, discussion of group identity has tended in recent years to revolve around the ways in which groups shape the practices, attitudes, and values of their members and the ways in which such membership leads to their political and social marginalization. The group and its identity is defined or characterized in terms of, say, ethnicity, gender, sexual orientation, religion, language, or region. Indigenous peoples in North America or Australasia have, for example, been excluded (both de facto and historically de jure) from or forced to assimilate into the political and economic mainstream because their groups were regarded as excluded from or outside of the norms of full or proper citizenship. The relationship between political and social norms defined in relation to a dominant group and other groups not conforming to those features or practices underpins the rise in real-world political discourse, sociological analysis, and philosophical thought of "identity politics," or "the politics of difference."

This discussion of the significance of group identity, though, suffers from an omission, which has been characteristic of much of the social-scientific literature, which fails to consider and spell out the *ontological* commitments of social-scientific discourse. In talking of the effects of group membership, the claim for group rights, the role of a group in bringing about some event or state of affairs, or the demand of one group to be compensated for the wrongs inflicted by another, social scientists and philosophers are brought to an old question: *What is a social group?* Hard on its heels come others: How can we individuate one group from another, and what makes a social group the very same one through time? Of course, these questions connect the philosophy of social sciences to more general issues

in *metaphysics* concerned with the individuation of objects and their diachronic identity conditions.

For the social sciences these old questions might now be regarded as superannuated, coming to be superseded by, inter alia, interests in social properties, collective intentional states, and systemic approaches to society. Notwithstanding the interest and importance of such issues, the old—ontological—questions have never gone away, for they are at the core of an understanding of the nature of the social sciences. The truth conditions of many claims about the social world depend upon the existential or referential status of groups such as nations, peoples, classes, communities, teams, tribes, and families. A proper understanding of what is said—of what we mean—turns on how we are to treat references to social groups. The justification of moral evaluations, the articulation of practical judgments and actions, and the formation of policies depend upon the objects of such judgments or actions being appropriate ones. In particular, it must be the kind of thing capable of sustaining such judgments and of being responsive to particular policies and actions.

The majority view among philosophers is that these questions can be answered in a very straightforward fashion. Groups can (at least in principle) be excluded from social-scientific discourse. Labeling a diverse set of views as *individualism,* this position maintains that individuals and their relations enjoy ontological and explanatory priority. Depending on its form, individualism holds groups to be identical to sets (or mereological sums of individuals or person stages), mere fictions or reductively analyzed out of social-scientific discourse.

The truths about groups are held to be expressible, without loss, as truths about *individuals.* The very notion of group identity becomes shorthand for the ways in which individuals relate and, typically through standing in certain kinds of relationships, conceive of themselves.

Others have argued that social groups cannot be identified with sets or aggregates or reductively analyzed out of our social-scientific descriptions, explanations, and predictions. The realist thesis holds groups to be composite material particulars capable of standing in causal and explanatory relations. A variant of a realist thesis disavows the materiality of social groups. Social substances, such as countries or social institutions, are taken to be spatiotemporally locatable but nonmaterial entities. But such entities are puzzling.

For (materialist) realism, then, at a high level of taxonomic categorization, groups feature alongside kinds such as organisms and artifacts. This entitlement to individuate groups as material objects relies on two key claims: first, that reference to groups in social-scientific and everyday discourse is ineliminable and, second, that ineliminability from our best theoretical model is the hallmark of realism in general. Granted these claims, realism cashes out group identity as a thesis for the individuation of a type of entity. It is now—the realist may insist—the job of the social sciences to elucidate why particular groups have certain features and to trace the ways in which different types of groups are treated and the impact of such treatment on their members.

Realism about the identity of groups brings with it the burden of addressing two challenges. First, the possibility of synchronic coextensive memberships of groups (e.g., the philosophy department and the wine appreciation society) suggests that two material particulars of the same kind can be in the same place at once. This appears to fly in the face of our commonsense understanding of material particulars. Second, the realist owes an account of the survival conditions of a group. A group can undergo change through time while remaining the same group, most obviously through changes in its membership. For most objects, the very same thing survives change in its parts, provided its parts continue to be organized through time in the form characteristic of that kind of object. Groups, though, by being composed of intention-forming persons, seem more prone to mergers, divisions, and changes in their defining characteristics (e.g., the prevalent attitudes and values) than artifacts or organisms. To hold that group identity is more than a way of talking about individuals in relations and refers to the identity of a group as such through time, these issues must be addressed by the realist.

In concluding, it is important to note that the issue of the ontological status of groups runs through or cuts across the question of whether a proper understanding of the social sciences is naturalistic, interpretive, critical, or postmodern. No position on the nature of the social sciences can ignore the further question.

Paul Sheehy

See also Collective Agents; Group Beliefs; Group Mind; Holism, in the Social Sciences; Individualism, Methodological; Realism and Anti-Realism in the Social Sciences; Social Ontology, Recent Theories of; We-Mode, Tuomela's Theory of

Further Readings

Effingham, N. (2010). The metaphysics of groups. *Philosophical Studies, 149,* 251–267.
Gilbert, M. (1989). *On social facts.* Princeton, NJ: Princeton University Press.
Gilbert, M. (1996). *Living together.* Lanham, MD: Rowman & Littlefield.
May, L. (1987). *The morality of groups.* Notre Dame, IN: University of Notre Dame Press.
Ruben, D.-H. (1985). *The metaphysics of the social world.* London, England: Routledge & Kegan Paul.
Sheehy, P. (2006a). *The reality of social groups.* Aldershot, England: Ashgate.
Sheehy, P. (2006b). Sharing space: The synchronic identity of social groups. *Philosophy of the Social Sciences, 36*(2), 1–18.
Tuomela, R. (1995). *The importance of us.* Stanford, CA: Stanford University Press.
Watkins, J. (1955). Methodological individualism: A reply. *Philosophy of Science, 22,* 58–62.

Group Mind

Talk of group minds has arisen in a number of distinct traditions, such as in sociological thinking about the "madness of crowds" in the 19th century and, more recently, in making sense of the collective intelligence of social insects, such as bees and ants. In this entry, we provide an analytic framework for understanding a range of contemporary appeals to group minds and cognate notions, such as collective agency, shared intentionality, socially distributed cognition, transactive memory systems, and group-level cognitive adaptations.

Introduction

While individual agents—for example, people, other living things, robots, and computer systems—are the most commonly invoked cognitive agents, the positing of group minds has a rich and diverse history in the cognitive, biological, and social sciences. Much of this history threads its way into contemporary discussions of group minds and closely related notions. For example, in envisaging a universal knowledge network that would serve to advance human civilization, 18th-century French encyclopedists, such as Denis Diderot and Jean D'Alembert, provided an intellectual anchor for contemporary thought about the Internet and the World Wide Web functioning as the hardware for a "global brain"—Francis Heylighen being among those who contributed to the latter notion. The evolutionary biologist David Sloan Wilson has drawn explicitly on 19th-century work on group minds in arguing that group-level cognitive adaptations could evolve by group selection. And Irving Janis's cautions about "groupthink" and its negative consequences echo the historically popular work of Gustav LeBon and others in the 19th century warning of the dangers of fermenting conditions that thwart individual rationality and behavior.

While there is much more to be said about the history of such appeals to group minds and their relationship to ongoing discussions, this entry focuses on the contemporary discussions themselves. In particular, it provides an overview of the major strands of recent work structured around what is *shared* within a group of individuals usefully thought of as constituting or producing some kind of group mind.

Collective Cognition and Group Minds

Contemporary proponents of the idea that groups can be the subjects of psychological properties in their own right often express their views without using the group mind idiom. This is at least partly because the vernacular concept of mind is closely intertwined with consciousness and a first-person awareness of our inner mental lives. Thus, speaking of "group minds" may seem to commit one to an implausibly strong ontology, one blurring the distinction between science and cybernetic fantasies about the technologically driven emergence of collective forms of consciousness, as Heylighen shows. Yet minds that stop short of having the full range of properties that our minds have are commonplace. Newborn human infants, nonhuman animals, and certain kinds of machines are recognized as possessing such minds, manifesting only some of the psychological states or abilities characteristic of the minds of normal adult human beings.

Contemporary proponents of collective cognition typically restrict their claims to particular kinds of psychological predicates taken to be shared by individuals and groups. Such predicates can be drawn from folk psychology (e.g., belief, intention, rational agency), may refer to classical mental faculties (e.g., memory, decision making, or problem solving), or involve more theoretically driven notions (e.g., adaptive information processing).

Such an approach to group minds retains the core of the older emergentist idea that a group as a

whole can have cognitive properties that none of its members has, properties that are irreducible to the properties had by those members, as for example, Georg Theiner and colleagues show. Group cognition is not simply the unstructured aggregation of individual cognition but the collaborative outcome of a division of cognitive labor among cognitive agents. The social interactions between individual agents can have unintended cognitive effects at the group level, similar to the self-regulating "invisible hand" of the marketplace. Analyzing social processes in terms of cognitive functions such as memory or problem solving serves to highlight certain important information-processing patterns that can also be observed when the same function is performed by individuals.

Three Kinds of Group Cognition

To further understand contemporary work relating to group minds, we focus on what it means to say that some psychological property or process is collectively possessed, enacted, or otherwise *shared* by a group of individuals. We distinguish three different interpretations of this claim (ranked in the order of ascending strength): coordinated cognition, collaborative cognition, and joint cognition.

Coordinated Cognition

The first and weakest interpretation corresponds to the sense in which two or more people can share tokens of the same type of mental state or attitude as a result of being members of the same group. For instance, the memories of individuals that are part of long-standing groups, such as religious groups, nations, or social classes, tend to converge toward a shared stable rendering of the past that actively shapes, and is also continually shaped by, the community's collective identity. The critical role of social interactions here is to provide a mechanism of cognitive alignment whose function is to enhance some (frequently nonpsychological) capacity of the group. Accordingly, the central feature of shared cognition in this first sense is the homogeneity or at least substantial similarity of individual members' perspectives. We suspect that many if not all appeals to group minds of this kind can be reduced to claims about individual minds, together with accounts of the social and other mechanisms coordinating the psychological states of those individuals.

Collaborative Cognition

A second, and stronger, way in which cognition can be collectively shared that is less readily reducible in this sense is exemplified by the way in which a group can share a cognitive workload by apportioning that workload among its members. Here, the emphasis rests on the integration of distributed cognitive resources within a group by various social and cognitive mechanisms that pool complementary pieces of information from its members. A particularly striking illustration of the cognitive virtues of this form of shared cognition is the "wisdom of crowds" effect, in James Surowiecki's phrase, which refers to the ability of intellectually diverse collections of independently acting individuals to make better decisions or predictions than those of its best members, provided that there is an efficient mechanism for aggregating individual knowledge that is immune to the dysfunctional influences of social conformism. Prediction markets (e.g., Intrade.com) are a successful real-world application of this phenomenon.

However, many complex cognitive activities that are collectively performed by groups are achieved in an even more collaborative fashion, involving levels of cognitive interdependence that require much more intimate forms of cooperation than those that are necessary to participate in a decentralized market mechanism. Much of contemporary work on group minds falls under this heading, often focusing on the performance of small groups concerned with highly structured cognitive tasks that lend themselves to a clear differentiation of functional roles. Consider three prominent examples.

Daniel Wegner coined the notion of a *transactive memory system* to express the idea that people who stand in continuing close relationships (e.g., as couples, families, coworkers) may develop a kind of supra-individual memory system. In principle, a transactive memory system is capable of remembering more than the sum of its members. However, this potential assembly bonus is often difficult to realize in practice, because members often do not know whom to ask, how to ask, and how to cue each other effectively. Only groups whose members trust one another's expertise, have a shared awareness of how encoding and storage responsibilities are apportioned, and have developed efficient practices of retrieval coordination are capable of producing emergent group-level memories.

The framework of *distributed cognition* differs from other collaborative approaches to collective information processing by placing a strong theoretical emphasis on how the material structure of the collaborative work environment, such as workspace design, tool use, or the physical format of verbal or written representations, sculpts and facilitates the social distribution of cognitive workload in space and time. For example, Edwin Hutchins (1995) concludes his celebrated computational analysis of coastal ship navigation aboard a Navy cruiser with the observation that "organized groups may have cognitive properties that differ from those of the individuals who constitute the group. These differences arise from both the effects of interactions with technology and the effects of a social distribution of cognitive labor" (p. 228). By considering the extended mechanisms that enable the real-time coordination of heterogeneous elements in support of various cognitive functions, rather than the spatial boundary between what's inside the head and what's not, proponents of the distributed-cognition approach aim to break down the traditional categorical divide between cognitive ("inner") and social-cultural ("outer") factors.

A third example of what we are calling collaborative cognition is drawn from the evolutionary literature on group selection. David Sloan Wilson has defended the view that whenever between-group selection dominates within-group selection, there are evolutionary pressures toward modes of social organization in which social groups begin to function like higher-level organisms. According to Wilson, the occurrence of such a major evolutionary transition during early human evolution has had a profound impact on our psychological makeup, such as the evolution of altruism, cooperation, religion, and morality. With an eye on tightly knit religious communities as a prime historical example, Wilson (1997) has argued that group selection can be strong enough to create situations in which "individuals might cease to function as independent decision makers and become part of a group-level cognitive structure in which the tasks of generating, evaluating, and choosing between alternatives are distributed among the members of the group" (p. 358).

Joint Cognition

The third, and strongest, sense in which cognition can be collectively shared requires that two or more people are jointly committed to assume particular mental states as a collective body, that is, to act as a plural subject of intentions, beliefs, or knowledge. For Margaret Gilbert, the formation of a joint commitment forges a special kind of nonsummative unity that is not an aggregate of individual commitments. Nor, she argues, does it, properly speaking, have any parts but instead brings into existence what Gilbert calls a "pool of wills" to which all group members are bound simultaneously and interdependently. Joint commitments in this sense give rise to distinctive group obligations and entitlements to which its members ought to adhere, including epistemic norms to which plural subjects can be held accountable. This creates the potential for genuine conflicts between individual and group rationality. As Christian List and Philip Pettit argue, group agents who have the capacity to recognize these obligations, and respond to them in a manner characteristic of reasoning subjects, constitute real, albeit nonnatural, persons.

Robert A. Wilson and Georg Theiner

See also Collective Agents; Collective Goals; Collective Intentionality; Collective Memory; Collective Moral Responsibility; Collective Rationality; Collective Values; Distributed Cognition and Extended-Mind Theory; Emergence and Social Collectivism; Group Beliefs; Plural Subjects

Further Readings

Gilbert, M. (1989). *On social facts*. London, England: Routledge.

Heylighen, F. (2011). Conceptions of a global brain: An historical review. In L. E. Grinin, R. L. Carneiro, A. V. Korotayev, & F. Spier (Eds.), *Evolution: Cosmic, biological, and social* (pp. 274–289). Volgograd, Russia: Uchitel.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge: MIT Press.

Janis, I. (1982). *Groupthink* (2nd ed.). Boston, MA: Wadsworth.

Le Bon, G. (1895). *The crowd: A study of the popular mind*. Dunwoody, GA: Berg.

List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford, England: Oxford University Press.

Surowiecki, J. (2004). *The wisdom of crowds*. New York, NY: Random House.

Theiner, G. (2011). *Res cogitans extensa: A philosophical defense of the extended mind thesis*. Frankfurt, Germany: Peter Lang.

Theiner, G., Allen, C., & Goldstone, R. (2010). Recognizing group cognition. *Cognitive Systems Research, 11,* 378–395.

Wegner, D. M. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). New York, NY: Springer.

Wilson, D. S. (1997). Incorporating group selection into the adaptationist program: A case study involving human decision making. In J. A. Simpson & D. T. Kenrick (Eds.), *Evolutionary social psychology* (pp. 345–386). Mahwah, NJ: Lawrence Erlbaum.

Wilson, D. S. (2002). *Darwin's cathedral*. Chicago, IL: University of Chicago Press.

Wilson, R. (2004). *Boundaries of the mind: The individual in the fragile sciences: Cognition*. Cambridge, England: Cambridge University Press.

H

Habitus

This entry traces the development of the concept of habitus in philosophy and sociology and focuses on Pierre Bourdieu's systematic theory of it, especially as a theory of action.

Classical Sources of the Concept

Habitus is a concept originally introduced by Aristotle, reworked by Thomas Aquinas, and used sporadically and unsystematically by some 19th-century European social theorists. The notion was revived and recruited for understanding the practical embodied bases of action by the French sociologist Marcel Mauss and the French philosopher Maurice Merleau-Ponty. But it was only in the work of the French sociologist and anthropologist Pierre Bourdieu that the concept was reintroduced with a more systematic intent into social theory as a viable analytic tool for the job of accounting for the cognitive components of action.

In its initial Aristotelian formulation, the notion of habitus is captured in the idea of *hexis* (habitus is the usual Latin translation of this Greek word). This refers to the state of possessing (or "having," Latin *habere*) an acquired, trained disposition to engage in certain modes of activity when encountering particular objects or situations. For instance, the essential capacity to regularly engage in virtuous action was understood, in the context of Aristotelian ethics, to be the primary exemplification of *hexis*. Aquinas would refine the application of the concept to ethical reasoning in further specifying the nature and content of the moral virtues. In Aquinas's rendering, the full virtuous personality is one who has, through effort and training, cultivated the proficiency to act in the morally required manner without effort, that is, a person for whom moral behavior becomes *second nature*—as put forth originally by Aristotle.

The Central Problem in the Sociology of Knowledge

In the sociology of knowledge, an attempt is made to account for the social or external bases of thought. This has a long tradition in Western social theory, but the two main classical sources are the essay on *Primitive Classification* written by the French sociologist Émile Durkheim and his nephew Marcel Mauss in 1902 and the monograph translated into English as *Ideology and Utopia* (1936), authored by the German sociologist Karl Mannheim. The Durkheim and Mauss essay is known for its bold claim that the categorical relations with which persons organize their understanding of the world of things are modeled after the social relations that govern the classification of persons as members of distinct groups. Mannheim, for his part, emphasized how the agonistic context of intellectual fields molds the content of ideological systems. In both of these statements, *a link between social structure and structure of systems of thought* is made, but the mechanisms that mediate this linkage remain murky.

Habitus as a Solution to the Central Problem

It was in his attempt to rebuild and rework this intellectual tradition—leaving behind its most egregious distortions and oversimplifications—that Pierre Bourdieu first resorted to the notion of habitus. His initial source of inspiration was admittedly an odd one: a book written by the art historian Irwin Panofsky (*Gothic Architecture and Scholasticism*) arguing how the aesthetic rules of composition of Gothic architecture—most clearly appreciated in the construction of the Gothic cathedral—were isomorphic to the habits of thought and learning inculcated in the medieval scholastic institution—most clearly appreciated in the construction of the *Summa,* the major scholarly treatise in which systems of ideas were expounded. In a postface written for the book in 1967, Bourdieu is struck by how in resorting to the Aristotelian notion of habitus to make sense of this correspondence, Panofsky is able to partially avoid the dilemma that plagued traditional sociology-of-knowledge proposals of a similar sort.

Habitus and the Theory of Action

The Connection Between Position and Disposition

Bourdieu would go on to systematize the idea of habitus for a more general understanding of the nature of social action in his subsequent work. In Bourdieu's rendering, which creatively melds a wide range of influences from Piagetian psychology, phenomenology, the history of epistemology, and onward to Lévi-Straussian structuralism, habitus is an *acquired system of schemes* that allows for everyday instances of perception, categorization, and the production of action and, most importantly, for the production of mundane judgments (e.g., judgments of moral propriety or impropriety, of likelihood or unlikelihood, or of certainty or uncertainty, or judgments of taste, e.g., likes and dislikes). The habitus endows action with a finality that is immanent or implicit in practice and that only appears *after the fact* as if it had been planned beforehand. The habitus, thus, subsumes the rationalistic theory of action inherited from utilitarian thought as an unnatural exception rather than the rule from which an understanding of action should be built.

In Bourdieu's rendering, the habitus is not just the producer of actions and reactions, but it is a *product* of the environmental conditions that the person encounters during ontogenetic development. In its essential status as a being open to (being modified by) the world, the person is bound to internalize the immanent necessities inscribed in that world in the form of habitus. As a form of internalized necessity, the habitus biases our implicit micro-anticipations of the kind of world that we will encounter at each moment, expecting the future to preserve the experiential correlations encountered in the past. This linkage of habitus to social position thus makes the concept of habitus inseparable from a theory of the differential distribution of social positions, which in Bourdieu's work takes the (natural) form of a field theory.

There has been some debate as to whether the habitus is a "collectivist" (holist) concept or is instead an "individualist" solution to the classic problem in the sociology of knowledge. The main point to note here is that in proposing that habitus emerges from experience, Bourdieu moves the debate beyond the traditional parameters of the sociology of knowledge because he locates the relevant realm upon which (whatever) regularities can be observed back into the (nonarbitrary) features of the world. In this respect, he moves beyond the impasse produced by the Durkheimian assumption of the arbitrary nature of the cognitive components of action and perception, which came to bedevil cultural anthropology.

Adaptation, Anticipation, and Change

Because the habitus is the product of adaptation to conditions, it is heavily predisposed to attempt to re-create the very same conditions under which the systems of skills and dispositions that it has most proficiently acquired can be most profitably put to use. The (tacit) recognition of this situation to skill match or mismatch is in effect the most obvious form of "practical reason" stored in the habitus. Conversely, the habitus tends to avoid those environments and situations for which it is not well adapted, in effect "refusing" that which is in fact objectively unavailable to it and liking or loving that which is already best fit to it. In this respect, the habitus is heavily weighted toward the past and biases choices in a way that lead to the conservation and constant reinforcement of already acquired capacities.

In its capacity as already accumulated (and thus "sunk-in") skill, the habitus carries with it a heavy load of inertia and only changes when external

conditions are so dramatically transformed as to permanently disrupt the capacity of the habitus to implement those strategies that worked in the past. This process of readaptation and reskilling is necessarily disruptive, and to some extent traumatic. Insofar as the habitus encompasses that which is most essentially a person's self, any rejection or transformation of the things that we do as second nature is in effect a rejection or a transformation of what a person "is" in the most fundamental sense.

Omar Lizardo

See also Embodied Cognition; Folk Psychology; Knowing-How Versus Knowing-That; Rule Following; Sociology of Knowledge and Science; Tacit Knowledge

Further Readings

Bourdieu, P. (1985). The genesis of the concepts of "habitus" and "field." *Sociocriticism, 1,* 11–24.

Bourdieu, P. (1990). *The logic of practice.* Cambridge, England: Polity Press.

Bourdieu, P. (2005). Postface to Erwin Panofsky, Gothic artchitecture and scholasticism. In B. W. Holsinger (Ed.), *The premodern condition: Medievalism and the making of theory* (pp. 221–242). Chicago, IL: University of Chicago Press.

Durkheim, E., & Mauss, M. (1963). *Primitive classification* (R. Needham, Trans.). Chicago, IL: University of Chicago Press. (Original work published 1902)

Green, A. I. (2008). Erotic habitus: Toward a sociology of desire. *Theory and Society, 37,* 597–626.

Lau, R. W. K. (2004). Habitus and the practical logic of practice: An interpretation. *Sociology, 38,* 369–387.

Lizardo, O. (2004). The cognitive origins of Bourdieu's habitus. *Journal for the Theory of Social Behaviour, 34,* 375–401.

Mannheim, K. (1936). *Ideology and utopia.* London, England: Routledge.

Hayek and the "Use of Knowledge in Society"

This entry explicates Friedrich Hayek's (1899–1992) social epistemology, an epistemology that gives due consideration to both the workings of the individual mind *and* the mechanics of the ambient sociality in which mind is enmeshed. On Hayek's account, mind and sociality are coevolved connectionist-like systems, the latter scaffolding the inherently constrained mind, thereby significantly reducing the epistemic transaction costs involved in the harvesting of knowledge.

Hayek's Basic Insight

Hayek's most abiding philosophical insight is the idea that "perfect" knowledge is unnecessary, impracticable, and irrelevant, and indeed for these very reasons, its indiscriminating pursuit can be pernicious. Hayek's specific targets were two species of "rationalism": (1) central planning (favored by collectivism) and (2) the abstract individualism of homo economicus (favored by orthodox economics). According to Hayek, these rationalisms fail miserably to appreciate that cultural complexity offers both the *fabric of possibility* and the *fabric of inherent constraint.* If one understands sociality to be a complex and necessarily *dynamic* weave of innumerable "spontaneous orders," then on Hayek's account, knowledge, paradoxically, can become less incomplete only if it becomes more dispersed.

Hayek's social epistemology is spread across his work and is not coextensive with the essay "The Use of Knowledge in Society," which appears in this entry's title. Hayek's core presupposition—that mind is subject to terminal cognitive and epistemological constraint—crucially informs his social epistemology, philosophy of social science, and social theory. Indeed, Hayek made a very distinctive link between mind and freedom: All institutions of freedom are essentially adaptations to the fundamental fact of cognitive constraint. Hayek's philosophical psychology is set out in its most extended and technical form in the much-neglected *The Sensory Order,* but its concerns are to be found in attenuated form throughout his substantial corpus.

Cognition and Sociality

"Cognitive constraint" connotes the idea that the human mind is terminally subject to cognitive and epistemological opaqueness, not only concerning its own mechanics (the mind–body problem) but also in terms of collective intentionality. The *fundamentum* and the *residua* of practical reasoning of multitudes condense into customs, practices, and traditions,

entailing that mind is inescapably culturally saturated: Knowledge is modulated and particularized. Cognition, is therefore, always epistemically perspectival despite a universally shared cognitive architecture. In order to "harvest" knowledge, Hayek postulates a *dynamic,* connectionist-like theory of mind (the nervous system characteristically being a universal pattern–seeking mechanism).

In many respects, for Hayek, mind is the analog of his connectionist-like theory of sociality; that is, both mind and sociality are conceived as classic instantiations of mutually *coevolved* spontaneous orders. This idea is absolutely critical to understanding Hayek. In much the same way that synapses are strengthened while unused linkages weaken and wither away, so too are paths to salient social knowledge strengthened or weakened—*social* connectionism, if you will.

Action and Social Knowledge

As a guide to action, social knowledge (know-how) cannot be reduced, abridged, or restated propositionally (as knowledge-that) without remainder. It is for this reason that the demand for a demonstrative and deliberate use of *reason* in matters of sociality will defeat the cognitive capacity of any one individual or too heterogeneous a group of individuals. Hayek at different times in his career variously attributed this inappropriate rationalism to certain economists, central planners, "scientistic" social theorists, and ideologues of both the Left and the Right. That knowledge cannot be aggregated, centralized, or otherwise fully harvested is, in Hayek's view, a misplaced pessimism. On the contrary, this state of affairs facilitates "computational" and cognitive efficiencies by enabling individuals to exploit environmental and social resources rather than having to encode everything relevant within the confines of the brain. It is this "situated" perspective that motivates Hayek's social externalism. For Hayek, this is the mark of advanced cognition—humankind's evolutionary propensity to diffuse propositional and practical knowledge or wisdom through external epistemic structures.

Communicative Systems of Knowledge

For Hayek the conditions for epistemic veracity are best fostered through the manifold of *spontaneous orders* or "communications systems" that compose the fabric of the civil (liberal) condition. Hayek was centrally concerned with the *communicative* aspect of knowledge—the acquisition, mediation, and transmission of (for the most part, third-party) knowledge in complex communities of knowers. A healthily functioning communications mechanism promotes computational efficiencies, a coordination mechanism for diverse wants, preferences, and interests. It should be noted that at no time did Hayek give ontological priority to the market (attributions of economism to Hayek are thoroughly misplaced). Whatever the market's epistemic strengths (and they are considerable), for Hayek the healthy functioning of a market presupposes other institutions and spontaneous orders. Indeed, a manifold network of spontaneous orders (science and democracy being two other prominent orders) best promotes the conditions for moral and political freedom or autonomy. No one order should subsume or impinge upon another: The exercise of epistemic immodesty within and between orders is bound to impoverish the liberal condition. It is the sine qua non of the liberal condition that there will always be inherent tensions—to make one order answerable to another order's metric is both rationalistic and, indeed, antiliberal.

Concluding Remarks

Hayek's intellectual fortunes have ebbed and flowed: his popularity with the public counterposed by his lack of popularity within the academy, his reputation being tarnished from having been sequestered by—and consequently unfairly associated with—the laissez-faire community. Hayek's appeal is now significantly broader and his prescience in matters epistemological and in the philosophy of mind is now much more richly appreciated. The lived subjectivity emphasized by the Austrian tradition, the Scottish Enlightenment's emphasis on emergent properties, the "situated" non-Cartesian wing of cognitive science, analytical social epistemology, and network theory—all find a confluence of interest in Hayek's work. If there were a slogan that captures Hayek's lifelong project, it is that Hayek socializes the mind and "cognitivizes" social theory.

Leslie Marsh

See also Austrian Economics; Complexity and the Social Sciences; Cooperation/Coordination; Distributed Cognition and Extended-Mind Theory; Invisible Hand

Explanations; Scottish Enlightenment: Influence on the Social Sciences; Situated Cognition; Spontaneous Order; Tacit Knowledge

Further Readings

Butos, W. N. (Ed.). (2010). The social science of Hayek's "The Sensory Order." *Advances in Austrian Economics, 13*. doi:10.1108/S1529–2134(2010)0000013023

Feser, E. (Ed.). (2006). *The Cambridge companion to Hayek*. Cambridge, England: Cambridge University Press.

Hayek, F. A. (1948). *Individualism and the economic order.* Chicago, IL: University of Chicago Press.

Hayek, F. A. (1952a). *The counter-revolution of science: Studies on the abuse of reason.* Indianapolis, IN: Liberty Fund Press.

Hayek, F. A. (1952b). *The sensory order: An inquiry into the foundations of theoretical psychology.* Chicago, IL: University of Chicago Press.

Hayek, F. A. (1960). *The constitution of liberty.* Chicago, IL: University of Chicago Press.

Hayek, F. A. (1978). *New studies in philosophy, politics, economics and the history of ideas.* Chicago, IL: University of Chicago Press.

Hayek, F. A. (1979). *Law, legislation and liberty: Vol. 3. The political order of a free people.* Chicago, IL: University of Chicago Press.

Hegelianism and Contemporary Epistemology

This entry on Hegelianism and contemporary epistemology construes the latter as analytic philosophy. In recent years, as analytic figures turn increasingly to Georg Wilhelm Friedrich Hegel, Hegelianism has become one of the strands in contemporary epistemology. Analytic thinkers tend to read on the basis of prior analytic commitments. The complex relation between analytic philosophy and Hegelianism has evolved greatly over the past century. The entry starts by going over the earlier historical ground that led to the recent resurgence of Hegelianism in the hands of analytical philosophers. The crucial primary element in this has to do with the rejection of idealism by early analytic philosophy and its subsequent rereading of idealism's patriarch, Kant, by later representatives of the analytic tradition.

The analytic approach to Hegelianism is influenced by factors such as the general analytic reception of idealism, Wittgenstein's later attack on empiricism, and the analytic reading of Kant. The general analytic reception of idealism resembles the Marxist reading of idealism. Marxism has always defined itself as the negation of Hegelianism, and Hegel is routinely understood as a German idealist. The Marxist turn away from Hegel was reprised by analytic philosophy as it was emerging in England around the turn of the 20th century. Bertrand Russell and G. E. Moore, the founders of analytic philosophy, both wrote dissertations on Kant, Russell on his views of mathematics, and Moore on his views of ethics. Both for a short time considered themselves to be idealists, and both later turned against it. Kant famously added a "Refutation of Idealism" to the second edition of his classic, the *Critique of Pure Reason,* to distance himself from "bad" idealism. Moore attacked idealism of all kinds, which, he claimed, denied the existence of the external world. Moore's attack resulted in the analytic interdiction of idealism, which has never been lifted.

A specifically analytic reading of Kant emerged in the writings of Anglophone philosophers like Graham Bird, Jonathan Bennett, Peter Strawson, and others. Strawson presented a reading of Kant shorn of transcendental idealism. Kant, of course, rejected naturalism in adopting transcendental idealism to respond to Hume. Strawson's naturalistic approach to Kant read the critical philosophy as if it were an early form of analytic empiricism.

Early analytic philosophy was thoroughly empirical. Russell featured various logical forms of empiricism, while Moore espoused both commonsensism and sense-data theory. In the meantime, Wittgenstein launched an attack on Moore in favoring a kind of contextualism.

The analytic attack on idealism turned attention away from Hegel, which was turned back to Hegelianism by the later Wittgenstein's attack on empiricism and the emergence of a specifically naturalistic, nonidealistic approach to Kant. The analytic turn or return to Hegel was further strengthened by Wilfrid Sellars in attacking the notion of givenness under the heading of *méditations hégéliennes.*

Building on the later Wittgenstein, Sellars clearly signals his receptivity to Hegel in his frontal attack on classical empiricism, which he calls the myth of the given. He borrows Hegelian arguments in turning

Hegel's critique of sense certainty against empiricism. If the immediate given cannot be known, then empiricism as it has been understood in the English tradition and perhaps even in Kant, if Kant subscribes to a successor form of English empiricism, is indefensible. In restating Hegel's argument in an analytical idiom, Sellars rejects the idea of immediate givenness, or in Hegelian terms *immediacy,* as no more than a myth in favor of the justification of claims to know within the so-called logical space of reasons. Although the critique of traditional empiricism is closely Hegelian, numerous other features in Sellars's theories, such as a scientism of sorts resulting from his appeal to the so-called space of reasons, are non-Hegelian, even anti-Hegelian.

Sellars's intervention in the debate helped reverse Hegel's bad analytic reputation while deepening the later analytic revolt against empiricism. His Hegelian approach to immediacy facilitated a modest analytic rapprochement with Hegel, yet always with Hegel shorn of his idealism. After Gottlob Frege, the problem of semantic reference became the main theme in analytic philosophy. Analytic figures attracted to Hegel were often among those who turned away from the traditional analytic concern with semantic reference. They include contemporary philosophers such as the late Richard Rorty, always a maverick among analytic writers; his former students Robert Brandom and John McDowell, who both approached Hegel through Sellars; Pirmin Stekeler-Weithofer, a German philosopher; the Australian Paul Redding; and others.

Analytic readings of Hegel are not always closely tied to the texts. Rorty, who describes Hegel as one of his main influences, advances an idiosyncratic view of the German thinker. In a well-known account of the pragmatist philosopher John Dewey, Rorty depicts Hegel as defending a historicism that amounts to the claim that we will not be destroyed by a giant comet.

Turning to the contemporary scene, Brandom and McDowell together constitute the Pittsburgh school of analytic Hegelianism. They have sharply divergent views of Sellars and of Hegel. Among current analytic philosophers, Brandom has most closely identified his position with pragmatism and Hegel. Brandom does not distinguish between idealism and a neo-analytic pragmatism. Hence, his reading of Hegel requires the most careful scrutiny; his pragmatic interpretation of Hegel is firmly intertwined with, even inseparable from, his interest in what he calls "inferentialism." Brandom has attracted attention for his suggestion that inferentialism, his name for a restatement of the familiar idea of inference to the best explanation, is anticipated by Hegel. Brandom's inferentialism is associated in his mind with a variation on traditional analytical metaphysical realism. His metaphysical realist claim that reality makes our views, say, of electrons or aromatic compounds, true or false is apparently incompatible with Hegel's empirical realism. McDowell is critical of both Rorty and Brandom. Unlike many other analytic figures, for instance, Brandom, who reads Hegel through Sellars, McDowell reads Hegel directly through the texts as well as through such Hegel scholars as Robert Pippin. McDowell avoids conflating Hegel with metaphysical realism. He contends that Hegel does not fall behind, but rather goes beyond, Kant in developing an approach to knowledge that, unlike Sellars's "space of reasons," clearly preserves an empirical constraint. This observation suggests that analytic figures interested in Hegel should breach the interdiction concerning idealism established by Moore more than a century ago. In a recent work, Paul Redding points to continental idealism, including Hegel, as potentially a *third way* between scientific naturalism, or analytic philosophy, and orthodox theism.

Tom Rockmore

See also Empiricism; Epistemology; Given, Myth of the; Idealism; Inferentialism; Pragmatism

Further Readings

Ameriks, K., & Stolzenberg, J. (Eds.). (2005). Deutscher Idealismus und die Analytischer Philosophie der Gegenwart [German idealism and contemporary analytic philosophy]. In *Internationales Jahrbuch des Deutschen Idealismus* [International yearbook of German idealism] (No. 3). Berlin, Germany: W. de Gruyter.

Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment.* Cambridge, MA: Harvard University Press.

Brandom, R. (2000). *Articulating reasons: An introduction to inferentialism.* Cambridge, MA: Harvard University Press.

Brandom, R. (2011). *Perspectives on pragmatism: Classical, recent, and contemporary.* Cambridge, MA: Harvard University Press.

McDowell, J. (1996). *Mind and world.* Cambridge, MA: Harvard University Press. (Original work published 1994)

McDowell, J. (2003). Hegel and the myth of the given. In W. Welsch & K. Vieweg (Eds.), *Das interesse des denkens: Hegel aus heutiger sicht* [The interest of thought: Hegel from today's perspective] (pp. 75–88). Munich, Germany: Wilhelm Fink.

Moore, G. E. (1903). The refutation of idealism. *Mind, XII*(4), 433–453.

Redding, P. (2009). *Continental idealism: Leibniz to Nietzsche.* London, England: Routledge.

Rockmore, T. H. (2007). *Idealism and analytic philosophy.* New Haven, CT: Yale University Press.

Rorty, R. (Ed.). (1998). Dewey between Hegel and Darwin. In *Truth and progress: Philosophical-papers* (Vol. 3, pp. 290–306). New York, NY: Cambridge University Press.

Sellars, W. (1998). *Empiricism and the philosophy of mind.* Cambridge, MA: Harvard University Press.

Strawson, P. (1966). *The bounds of sense: An essay on Kant's critique of pure reason.* London, England: Methuen.

Wittgenstein, L. (1969). *Über Gewissheit* [On certainty] (G. E. M. Anscombe & G. H. von Wright, Eds.; D. Paul & G. E. M. Anscombe, Trans.). Oxford, England: Blackwell.

Herder's Philosophy of History

This entry provides an overview of Johann Gottfried Herder's (1744–1803) philosophical views on history and in particular his championing of "hermeneutic historicism," his rejection of the divide between nature and history or culture, and his critical stance vis-à-vis the Enlightenment. Herder's views have become important in recent discussions of the philosophy of history and, especially, in relation to what the Enlightenment is claimed to have bequeathed to the social sciences.

With his two major books, *Another Philosophy of History for the Education of Mankind* and *Ideas for a Philosophical History of Mankind,* Herder became a seminal figure in the emergence of *hermeneutic historicism*. Already in his essays on Shakespeare and Ossian for the famous *Sturm und Drang* manifesto of 1773, *On the German Manner and Art,* he had argued for the developmental character of language, literature, and society and for human situatedness in a historical-cultural context, the key tenets of hermeneutic historicism. For Herder, the task of historical reconstruction was always to explain an author in terms of his own time, not to transfigure him according to the taste of one's own century. History should define the triad of antecedents, the moment, and the posterity of an author: what the first offered him, what the second helped or hindered, and what the third carried further.

The main target for Herder's first book was the "philosophical history" articulated by Voltaire. The latter clearly believed European civilization, especially his own country France in the age of Louis XIV, to be at the forefront of progress in history, wherein reason brought about liberation from religious obscurantism and traditional domination. However, considerable ethnographic and historical evidence was emerging that contradicted the timelessness and ubiquity of European conceptions of reason, moral improvement, and political order. Variety in values across geographical and historical contexts challenged not only the doctrine of progress but also the very idea of the universality of reason. Herder championed this historicist counterpoint, and Voltaire's complacent Eurocentric presentism especially offended him. Already in 1769, Herder could comment ironically that his century was too refined, too political, and too philosophical to appreciate earlier epochs. By 1774, his outrage with smug European ethnocentrism and concomitantly brutal imperialism erupted in the scathing polemic of *Another Philosophy of History.* He lambasted the Enlightenment propensity to take the polish of its century for the *non plus ultra* of mankind. Earlier epochs and different cultures could not be read off simply as stages on the way to some later, higher order: Every nation had its own center of well-being within itself, just as every globe had its center of gravity. Herder found symbolic integrity and intelligibility in human artifacts where the Enlightenment could find only fantastic excess or benighted superstition. His conception of history as composed of individualities with their own intrinsic principles of unity, together with his method of

empathetic understanding without moral condescension, disputed the Enlightenment's privilege of the (European) present and doctrine of linear progress. While Herder still articulated a developmental scheme for world history, his metaphor of the phases of a life also gave distinctive value and merit to earlier phases of history and their expressive and cultural powers.

A decade passed before his second great work began to appear. The difference in tone and approach of the new work was striking. It was both more irenic and more synthetic. Crucially, Herder was now convinced that there could be no categorical divide between nature and history. Adopting Buffon's strategy of a "natural history of man," [*sic*] his grand project in the *Ideas for a Philosophical History of Mankind* was to find how man as a creature of nature figured in man as an artifice of culture. He proposed to combine cultural history with geography and natural history to create a natural history of peoples. Works of human purpose had a structure analogous to that of living organisms, for their various parts were coordinated within a whole whereby their particular nature and function became comprehensible. Not only might a poem or a genre be read in this new manner, but so could peoples, cultures, states, and epochs. Thus, Herder helped establish in German the terms for a distinctive "spirit" unique to a people, a nation, or an epoch: *Volksgeist, nationaler Geist, Zeitgeist*. This spirit informed every artifact of a culture, so that it could be read out of all forms of its practice—from folkways to political constitutions, from musical compositions to business contracts. But the highest and most revealing form would be the literary expression of that culture.

Herder believed in grasping "universal history," as it was called in the late 18th century. For him, the Enlightenment was wrong not in seeking such a comprehensive vision but in identifying the whole with the end or *telos*. He called for "history of mankind" not to trace the trajectory of "progress" but to discriminate the varieties of human excellence. Totality, for Herder, could only signify a historical ensemble of distinctive actualizations of human potential over the course of time—not in hermetic isolation, not without partial cumulation and mutual influence, but emphatically without a linear, progressive *telos*. Combining organicism with development, Herder could recognize higher-order meaning in history. This is the essential feature of his concept of *Humanität*. In *Ideas,* Herder wrote that this term encompassed everything he was trying to say concerning man's disposition to reason and liberty, to refinement and taste. In *Letters for the Advancement of Humanity,* he elaborated that the term evoked the character of our kind, but we were born with this character only as a disposition, which needed to be expressed concretely to be fulfilled.

Herder first effectively asserted the principle that history must judge epochs, peoples, or cultures according to their own intrinsic principles, not by some external standard imposed from the historian's own time. While, to be sure, Giambattista Vico had argued earlier for such interpretive principles, he was not read widely until much later; thus, it was with Herder that the historicist view entered decisively into the human sciences.

John H. Zammito

See also Enlightenment, Critique of; Hermeneutics, Phenomenology, and Meaning; Historicism; Philosophy of History; Vico's *Scienza Nuova*

Further Readings

Adler, H., & Koepke, W. (2009). *A companion to the works of Johann Gottfried Herder.* Rochester, NY: Camden House.

Herder, J. G. (1997). *On world history: An anthology* (H. Adler & E. Menze, Eds.). Armonk, NY: M. E. Sharpe.

Herder, J. G. (2004). *Another philosophy of history and selected political writings* (I. D. Evrigenis & D. Pellerin, Eds.). Indianapolis, IN: Hackett.

Zammito, J. H. (2002). *Kant, Herder and the birth of anthropology.* Chicago, IL: University of Chicago Press.

Zammito, J. H. (2009). Herder and historical metanarrative: What's philosophical about history? In H. Adler & W. Koepke (Eds.), *A companion to the works of Johann Gottfried Herder* (pp. 65–91). Rochester, NY: Camden House.

Hermeneutics, Phenomenology, and Meaning

Both hermeneutics and phenomenology have been influential in the development of certain schools of social science as well as in the philosophy of social sciences. Recently, rereadings of certain schools of

phenomenology by Anglo-American philosophers and European cognitive scientists have led to a renewed interest in phenomenology, especially in contemporary accounts of embodied cognition. This entry surveys the history and main tenets of hermeneutics and phenomenology, presents the protagonists, and explains the central place accorded to meaning.

Hermeneutics

Hermeneutics as the art of interpretation or understanding has a long history in theology, particularly in relation to the interpretation of *written* texts (especially the Bible), and in law with regard to the establishment of legal precedents, but in the 19th century, it was expanded into a general theory of human interpretative understanding. Aristotle's *On Interpretation* is often regarded as an early treatise on hermeneutics, and St. Augustine's *On Christian Doctrine* proposes some classic hermeneutical principles—for example, that the part must be read in relation to the spirit of the whole text. Hermeneutics was revised by theologians in the 19th century and was integrated into European philosophy in the 20th century by Wilhelm Dilthey, Martin Heidegger, Hans-Georg Gadamer, Paul Ricoeur, and others.

In the 19th century, hermeneutics was developed especially by Friedrich Schleiermacher (1768–1834). Schleiermacher wanted to recover the original meanings of the events of the New Testament with all layerings of subsequent tradition stripped away. This required attempting to reconstruct the original meanings of terms in ancient texts and reconstructing the original worldview of the participants. Hermeneutics, for Schleiermacher, proceeded through empathic intuition; it was necessary to put oneself in the shoes of one of the original, unlettered hearers of Jesus' parables, for instance. His assumption that there is a single underlying meaning (often called the "intent of the author") has been criticized by deconstructionists such as the 20th-century French philosopher Jacques Derrida. Wilhelm Dilthey (1833–1911) broadened hermeneutics to be a general theory of human experience in history. The German phenomenologist Martin Heidegger (1889–1976) was influenced by reading Dilthey and by his Marburg colleague, the Lutheran theologian and New Testament scholar Rudolf Bultmann (1884–1976). In *Being and Time* (1927), Heidegger claimed that phenomenology had to be pursued in a hermeneutic manner and proposed the new methodology of *hermeneutical phenomenology*. In that work, Heidegger also announces a procedure of historical "destruction" (*Abbau, Destruktion, Zerstörung*) of the history of philosophy, aimed at recovering the original sense of the fundamental concepts of the philosophical tradition—for example, recovering what the Ancient Greeks originally meant by terms such as *being* (*ousia*) or *nature* (*physis*) prior to their reification in the Western metaphysical tradition.

Hermeneutical phenomenology was subsequently developed in Germany by Heidegger's student Hans-Georg Gadamer (1900–2002) and in France by Paul Ricoeur (1913–2005). A basic tenet of hermeneutics is that all understanding involves interpretation on the basis of certain presuppositions, prejudgments, or prejudices. In this sense, Gadamer claims, all understanding involves misunderstanding. There is no neutral view from which to grasp meaning; we understand situations from the perspective that we occupy. Often, hermeneutics works on written texts, but Gadamer, following Heidegger, broadened hermeneutics to the whole project of human self-understanding. Human beings are finite beings, limited by their particular language, their education, and their standpoint in history. All understanding has to recognize the finitude of human participants. Humans are essentially involved in the historically situated and finite task of understanding the world, a world encountered and inhabited in and through language. Human existence is characterized by "linguisticality" (a term Gadamer takes from Schleiermacher). Gadamer speaks of hermeneutical understanding on the model of conducting a conversation. A genuine conversation will go to places that none of the participants anticipated. Understanding is something that *happens* to people rather than something they control. What is aimed at is mutual understanding through a "fusion of horizons."

Phenomenology

Phenomenology emerged at the end of the 19th century as a systematic methodology aiming at describing human experience and its objects precisely in the manner in which they are experienced, without applying any presuppositions and paying particular attention to the mode of givenness of experiences. Phenomenology is a discipline, therefore, that is

sensitive to the varieties of ways in which meaning presents itself to subjects open to this disclosure of meaning. Both hermeneutics and phenomenology are concerned with the nature of sense (*Sinn*) and meaning, or significance (*Bedeutung*). Both disciplines offer ways to disclose or uncover the meanings latent in texts or experiences, meanings that have become distorted or covered up due to the operation of prejudice, tradition, and general misunderstanding. Phenomenology attends to the way our experiences are actually given to us; it aims to recuperate the "givenness" (*Gegebenheit*—a term Heidegger says is the "magic word" of phenomenology) of experience and to resist reductionist efforts to construe experience in a naturalistic manner. Phenomenology pays attention to the nonlinguistic sense found, for instance, in human perception, whereas hermeneutics tends to focus on linguistic meaning.

Brentano and Husserl

Phenomenology was developed by Edmund Husserl (1859–1938), building on the descriptive psychology of his teacher Franz Brentano (1838–1917). Brentano's insight that mental acts were *intentional*—directed at an object—gave Husserl the inspiration to develop phenomenology as a method for describing the essences of experiences (perceptual, emotive, and cognitive) as well as their intended objects in a nonreductive way. Central to phenomenology is the idea that experience is intentional; it is directed at an object (Husserl speaks of *noema*—originally a Greek word meaning "meaning") and understands that object in a particular way (through a particular "mode of givenness"). Every experience intends an object in its own manner. Husserl speaks of this correlation between intention and the intended object as a *noetic-noematic* correlation. Husserlian phenomenology claims that entities in the world—and other people, animals, cultural products, and so on—are experienced in a particular way through the dominant "natural attitude." It requires a particular effort of suspension of belief, which Husserl calls the *epoché* (a Greek term that means "abstention" or "suspension"), in order to uncover the intentional achievements that underlie the world, as disclosed in the natural attitude. Husserl speaks of performing a "phenomenological-transcendental *epoché*" in order to lead back to the pure sources of intentional experience. Husserl then seeks to perform a number of "reductions," for example, the "eidetic reduction," in order to move from the individual experience to the essence (cf. the Greek word *eidos* for "essence") of that experience or its object. Phenomenology is, for Husserl, an a priori *eidetic description*. In order to arrive at essences, all belief in the actual world must be bracketed. Phenomenological insight—"eidetic viewing" (*Wesensschau*)—is achieved by carrying out a free eidetic variation to identify the invariant (*eidos*) and therefore necessary features of the object.

Husserlian phenomenology offers particularly powerful analyses of conscious experience, especially perception, memory, imagination, and judgment, as well as time consciousness and the experience of others in empathy (*Einfühlung*). Husserlian phenomenology also laid stress on embodiment and intersubjectivity. Traditional epistemology and philosophy of mind neglected the manner in which human beings are embodied.

In his later writings, Husserl concentrated on understanding how the sense of a common shared world is intentionally constituted. He contrasted the world as scientifically construed using formal methodologies with the "life-world" (*Lebenswelt*) in which human beings live and encounter each other most of the time. This led Husserl to a phenomenological critique of the manner in which modern scientific knowledge, as inaugurated by Galileo, had become distorted into a naturalistic ideology that reified objects as experienced "naively" in the natural attitude.

Hermeneutics and Phenomenology

Heidegger

Husserlian phenomenology exercises the *epoché* to lived experiences in order to reveal their meaning; hermeneutics, on the other hand, attempts to disclose meaning by liberating the experience from the historically transmitted tradition. Heidegger's introduction of hermeneutics into phenomenology was a way of neutralizing or at least exposing the operation of *prejudice* in our understanding. Prejudices for him cannot be eliminated, but at least they can be made transparent, acknowledged, and our corresponding insights put in correlation with these prejudgments so that our understanding progresses in a "circular" manner (Heidegger speaks

of the "hermeneutic circle"), going backward and forward between what is understood and the manner in which it is understood. To ask a question is already to anticipate in a certain sense what counts as an answer. There must be a fore-understanding to all understanding.

In *Being and Time* Heidegger proposes a radical description of human existence, which he prefers to call "existence'" (*Dasein*), since he regards terms like *consciousness* (favored by Husserl) to be too overlain with metaphysical presuppositions. Heidegger deliberately aims to make human existence unfamiliar by describing it in entirely novel terms. According to Heidegger, traditional philosophy since the Ancient Greeks has taken the nature of human existence more or less for granted. Human beings have been understood naturalistically since the time of Plato and Aristotle as "rational animals"; the religious traditions of the West, specifically Judaism and Christianity, have treated human beings as being somehow images of the divine nature and have sought to interpret human existence against the backdrop of the assumed eternal, unchanging existence of the divinity, in contrast with which human life is regarded as fleeting and inconsequential, a "vale of tears." Influenced by the idiosyncratic writings of the Christian existentialist Søren Kierkegaard (1813–1855), Heidegger wanted to revisit human existence and examine it in its concreteness, temporality, historicality, and finitude. To be human is to care about one's existence. Human existence is not something that simply "occurs," is "present-at-hand," is simply "there." Human existence is distinguished by the fact that individuals *care* about their lives; existence *matters* to the existing self. Human existence is individualized, and characterized by "mineness" (*Jemeinigkeit*), but at the same time human beings need familiarity, self-forgetfulness in the routines of the everyday, and understanding oneself as everyone else does, as "the one" (*das Man*). Human existence has a tendency to seek the familiar and the routine, what Heidegger calls "everydayness" (*Alltäglichkeit*). In this everydayness, time is experienced in a certain way that excludes the possibility of authentic selfhood, which Heidegger associates with free decision. The essence of human existence is, as Heidegger puts it, its "to-be." Humans are engaged in projects that cast them forward into the future; Heidegger speaks of the *ecstatic* (Greek *ekstasis,* "to stand outside") character of existence. Human beings are essentially temporal and finite and are essentially incomplete because of death. Human existence is characterized by "thrownness"; that is, humans find themselves in an always already constituted world of established meanings. In his later thought, Heidegger suggests that poetic creation offers a way in which language can become authentic, combating the "idle talk" (*Gerede*) of everyday inauthentic discourse.

Levinas

Influenced by Husserl and Heidegger, the Lithuanian-born Emmanuel Levinas (1906–1995), who became a French citizen, identifies a kind of immeasurable, even infinite, *desire* that drives human existence and goes far beyond the satisfaction of needs. Levinas is critical of the Western philosophical tradition for its pursuit of knowledge as a kind of domination over being, a will-to-power that undervalued the experience of recognition and respect for others. Levinas uses the term *face* to capture the uniqueness of our experience of the other. The face is something unique, irreplaceable, supremely individual, and expressive, and yet it is also vulnerable and, in a way, naked. The face presents the other in a very special way. Face-to-face relations with others are at the center of Levinas's phenomenology. The face of the other awakens a responsibility in me, and from that point of view, there is a kind of asymmetry in my obligation to the other person. From *my* perspective, I am more responsible than the other person. I can personally experience my own responsibility. Nevertheless, despite his criticism of the Western philosophical tradition, and despite his expressed wish to leave behind the "climate," as he put it, of Heidegger's philosophy, Levinas always presented himself as a disciple of Husserl and following in his tradition of phenomenology.

Sartre, Merleau-Ponty, and Ricoeur

After Husserl and Heidegger, phenomenology in France took an explicitly existentialist direction in the work of Jean-Paul Sartre (1905–1980) and Maurice Merleau-Ponty (1908–1961).

In his philosophical treatises and essays as well as in his novels and plays, Sartre provided brilliant phenomenological descriptions of human existence. Deeply influenced by his reading of Husserl and Heidegger (and, later, Karl Marx), Sartre expands

phenomenology in an existential direction, emphasizing human individuality, freedom, and finitude. His *Being and Nothingness,* published in 1943, is an essay in "phenomenological ontology." He characterizes human consciousness as a "negativity" or "nothingness" (*le néant*) that is the opposite of everything that has being (*l'être*). Consciousness is essentially intentional, it is directed at what it is not, and it is not any of the things it is conscious of. Human existence is essentially characterized by a kind of emptiness and void—this itself is the experience of freedom that people find frightening and dizzying (Sartre calls this experience "nausea," *la nausée*). Humans are not determined by some fixed set of characteristics (character) but are in fact always free—free to say no, free to resist. But existence itself is always contingent. Being just is; as such it is "superfluous" (*de trop*). There is no ultimate reason as to why things are the way they are. Things have sense and significance only insofar as they form a part of human projects. For Sartre, these "projects" come about by an act of fundamental decision or free choice. Ultimately, the human project is to be God, to be absolute master of one's actions, and also to be a complete being. This, however, is impossible. Human beings die. For Sartre, the experience of absolute freedom is so overwhelming that many people run away from it and try to shore up their lives with certainties, embracing social roles as if they were essential attributes of their being. This is what Sartre calls "bad faith" (*mauvaise foi*). In contrast, the authentic life is a life where one is always conscious of one's freedom to choose. In his later works, Sartre tried to integrate this existentialist account of the human condition into a more Marxist-inspired social philosophy.

In his major work, *Phenomenology of Perception* (1945), Merleau-Ponty offers a phenomenological account of human beings as embodied "being-in-the-world" (*être au monde*), a corrective to the one-sided accounts of experience found in what he calls generally "intellectualism" and "empiricism" (sometimes "sensationalism"). Merleau-Ponty's way of overcoming these oppositions is to focus on the complex and ambiguous nature of human embodiment or "incarnation." Phenomenology aims to disclose how we experience ourselves as embodied beings in the world. Merleau-Ponty stresses our implicit and unspoken knowledge of our own bodies. Merleau-Ponty wants to explore the complex ways our bodies relate to the world in "prereflective" lived, natural experience. He wanted to be present, as he put it, at the birth of our world. This interest led him in his later works to studies of infants (he succeeded the Swiss psychologist Jean Piaget), disabled people, and so on. Merleau-Ponty was particularly influenced by Husserl's exploration of embodiment in his *Ideas II* and by his later investigation of the pregiven life-world. Merleau-Ponty did not believe that this being-in-the-world can be uncovered by reflection alone. He sought to examine brain-damaged persons in whom the original assumed link with the world is broken, in order to display what is taken for granted in everyday experience. Contemporary cognitive scientists and philosophers have shown the importance of Merleau-Ponty's ideas (some directly inspired by him) in the growing interest in new theories of embodied cognition.

Finally, the French philosopher Paul Ricoeur has been quite influential in certain quarters in the philosophy of the social sciences on the Anglo-American side, introducing hermeneutics against what were considered positivist perspectives on social theory. Ricoeur was initially interested in the existentialist French philosopher Gabriel Marcel and the German philosopher Karl Jaspers, but he also read Husserl and Heidegger. His first publication was a study of Husserl's phenomenology, and in the same year, 1950, he published *Freedom and Nature: The Voluntary and the Involuntary,* the first volume of a phenomenology of the will. In this volume, he argued that human beings live at the intersection of the voluntary and involuntary. The second volume later appeared in two parts: *Fallible Man* and *The Symbolism of Evil.* In these works, Ricoeur moved to a new hermeneutical approach. For Ricoeur, an essential hermeneutical question is "Where are you coming from?" or, more literally, "From where are you speaking?" (*D'ou parlez vous?*). Phenomenology must involve a hermeneutical investigation of—or "detour" through—the symbolic domain of language and other forms of culture. For Ricoeur, we understand ourselves only through a long detour—through the way of symbols. Symbols give rise to and structure thought. Ricoeur went on to note that symbols are given radically conflictual readings from differing ideological standpoints. Always culturally situated, a hermeneutical phenomenology cannot finally resolve such conflicts. Its task is rather to uncover and delineate

the theoretical framework of each interpretative standpoint. In his terminology, a methodology of suspicion is to be coupled with one of affirmation. Ricoeur's hermeneutics always pays close attention to language, including the use of metaphor and the manner in which meaning is structured as a narrative. In his later work, he developed an account of the nature of the self involving the manner in which self always involves a relation to the other.

Conclusion

Phenomenology continues to have a strong influence because of its recognition of the first- and second-person perspectives that complement the objective, third-person approach found in the natural and social sciences. Phenomenology is particularly helpful in describing the life of embodied intentional consciousness, both singular and plural. Hermeneutics, and especially the emphasis that interpretation involves narrativity, continues to play a strong role in the cognitive and social sciences.

Dermot Moran

See also Being-in-the-World; Contemporary French Philosophy and the Social Sciences; Embodied Cognition; Empathy; Existential Phenomenology and the Social Sciences; Idealism; Intentionality; Intersubjectivity; Life-World; Neural Hermeneutics; Objectivity; Phenomenological Schools of Psychology; Scheler's Social Person; Social Construction of Reality

Further Readings

Brentano, F. (1995). *Psychology from the empirical standpoint* (A. C. Rancurello, D. B. Terrell, & L. McAlister, Trans., with a new preface by P. Simons). London, England: Routledge. (Original work published 1874)

Crowell, S. G. (2001). *Husserl, Heidegger, and the space of meaning*. Evanston, IL: Northwestern University Press.

Dreyfus, H. L. (1991). *Being in the world: Commentary on Heidegger's "Being and Time": Division 1*. Cambridge: MIT Press.

Gadamer, H.-G. (1977). *Philosophical hermeneutics* (D. E. Linge, Trans. & Ed.). Berkeley: University of California Press.

Gadamer, H.-G. (1989). *Truth and method* (2nd ed.; J. Weimsheimer & D. G. Marshall, Trans.). London, England: Sheed & Ward. (Original work published 1960)

Heidegger, M. (1962). *Being and time* (J. Macquarrie & E. Robinson, Trans.). Oxford, England: Blackwell. (Original work published 1927)

Husserl, E. (1960). *Cartesian meditations: An introduction to phenomenology* (D. Cairns, Trans.). The Hague, Netherlands: Martinus Nijhoff. (Original work published 1931)

Husserl, E. (1970). *The crisis of European sciences and transcendental phenomenology. An introduction to phenomenological philosophy* (D. Carr, Trans.). Evanston, IL: Northwestern University Press. (Original work published 1936)

Husserl, E. (1999). *Husserl collected works: Vol. 8. The idea of phenomenology* (L. Hardy, Trans.). Dordrecht, Netherlands: Kluwer.

Husserl, E. (2001). *Logical investigations*. New York, NY: Routledge. (J. N. Findlay, Trans., with a new introduction by D. Moran and a preface by M. Dummett; 2 vols.). (Original work published 1900–1901)

Levinas, E. (1969). *Totality and infinity* (A. Lingis, Trans.). Pittsburgh, PA: Duquesne University Press. (Original work published 1961)

Luft, S., & Overgaard, S. (Eds.). (2011). *The Routledge companion to phenomenology*. New York, NY: Routledge.

Merleau-Ponty, M. (1962). *Phenomenology of perception* (C. Smith, Trans.). London, England: Routledge & Kegan Paul. (Original work published 1945)

Merleau-Ponty, M. (1964a). *Sense and nonsense* (H. Dreyfus, Trans.). Evanston, IL: Northwestern University Press.

Merleau-Ponty, M. (1964b). *Signs* (R. McCleary, Trans.). Evanston, IL: Northwestern University Press.

Merleau-Ponty, M. (1968). *The visible and the invisible* (A. Lingis, Trans.). Evanston, IL: Northwestern University Press.

Moran, D. (2000). *Introduction to phenomenology*. London, England: Routledge.

Moran, D. (2005). *Edmund Husserl: Founder of phenomenology*. Cambridge, England: Polity Press.

Ricoeur, P. (1966). *Freedom and nature: The voluntary and the involuntary* (Erazim Kohak, Trans.). Evanston, IL: Northwestern University Press. (Original work published 1950)

Ricoeur, P. (1967). *The symbolism of evil* (E. Buchanan, Trans.). New York, NY: Harper & Row.

Ricoeur, P. (1974). *The conflict of interpretations: Essays in hermeneutics* (D. Ihde, Trans.). Evanston, IL: Northwestern University Press.

Ricoeur, P. (1981). *Hermeneutics and the human sciences* (J. B. Thompson, Trans.). Cambridge, England: Cambridge University Press.

Ricoeur, P. (1986). *Fallible man* (C. A. Kelbley, Trans.). New York, NY: Fordham University Press.
Sartre, J.-P. (1975). *Between existentialism and Marxism* (J. Mathews, Trans.). London, England: Pantheon.
Sartre, J.-P. (1995). *Being and nothingness: An essay on phenomenological ontology* (H. Barnes, Trans.). London, England: Routledge.
Sartre, J.-P. (2007). *Existentialism is a humanism* (C. Macomber, Trans.). New Haven, CT: Yale University Press.
Spiegelberg, H. (with K. Schuhmann). (1994). *The phenomenological movement: A historical introduction* (3rd ed.). Dordrecht, Netherlands: Kluwer.
Stein, E. (1989). *Collected works of Edith Stein: Vol. 3. On the problem of empathy* (W. Stein, Trans.). Washington, DC: ICS.
Zahavi, D. (2003). *Husserl's phenomenology.* Stanford, CA: Stanford University Press.

Heterodox Economics

Heterodox economics is now the fashionable expression describing the set of numerous nonorthodox schools of thought that have emerged through time. Heterodox economics is in opposition to orthodox economics, also known as mainstream economics or neoclassical economics. The purpose of this entry is to present the elements shared by the economists of the various heterodox schools of thought and to identify these schools.

Heterodox Schools in Economics

It is fruitful to start by making a distinction between dissenters in economics and heterodox economists. Dissent may express itself either within or outside the orthodoxy. Orthodox economics includes the mainstream—which roughly corresponds to the textbook view—and orthodox dissenters, who reject some assumptions of mainstream analysis, trying to push it to the frontier, while remaining within orthodoxy. By contrast, heterodox economists are dissenters who reject the validity of orthodox economics. While orthodox dissenters only deviate somewhat from the mainstream, heterodox dissenters are heretics.

There are a large number of heterodox schools of thought. While some of these distinctions reflect fundamental beliefs, they also occur as a result of field specialization. Obviously, Marxists, or radicals, are part of heterodox economics. So is Institutional economics, or at least Old Institutionalism. Post-Keynesian economics, associated with John Maynard Keynes, is usually subdivided into its Fundamentalist, Kaleckian, Kaldorian, and Sraffian (or Neo-Ricardian) branches and contains, with Neo-Marxism, the main heterodox contribution to macroeconomics. The French regulation school, with its U.S. version—the Social Structure of Accumulation school—is an attempt to achieve a grand synthesis by combining the Marxist, Institutionalist, and post-Keynesian insights, along with a historical analysis. Another important heterodox school is that of evolutionary economics, associated both with the Veblenian approach in institutional economics and with Schumpeterian economics, mostly concerned with innovations and their diffusion.

Other heterodox schools include the Structuralist school, mainly developed in Latin America; social economics and humanistic economics, which emphasize ethical norms, also tied to anti-utilitarism; the economics of conventions, which underline rules and accepted norms; feminist economics, with its focus on gender issues, most of which evades mainstream theory; ecological economics, which is the heterodox version of environmental economics; and behavioral economics, or more precisely the stream that studies economic psychology and behavior without attempting to repair mainstream rationality. There are many other heterodox schools or traditions, such as Buddhist economics, Gandhi economics, Gesell economics, and Henry George economics. As for Austrian economics, it is often included within heterodox economics, but as we shall see, this classification may not be appropriate.

Common Features of Heterodox Economics

Heterodox economists are more than mere critics of the mainstream. Besides the institutional and theoretical ties that unite the members of the various schools and the fact that the works of some authors are highlighted by several heterodox traditions, there are broad features that characterize heterodox schools in a positive and essential way—the *presuppositions* of their research programs.

Table 1 summarizes the methodological issues and beliefs that separate heterodox from orthodox economics. Heterodox economics is based on realism, reasonable rationality, some degree of holism, a

Table 1 Presuppositions of the Heterodox and Orthodox Research Programs

Presupposition	*Heterodox Schools*	*Orthodox Schools*
Epistemology/ ontology	Realism	Instrumentalism
Rationality	Reasonable rationality, satisficing agent	Hypermodel-consistent rationality, optimizing agent
Method	Organicism, holism	Individualism
Economic core	Production, growth	Exchange, allocation, scarcity
Political core	Regulated markets	Unfettered markets

Source: Author.

concern with production and growth instead of scarcity, and a belief in the need for regulated markets. Since economists of the Austrian school disagree with the last three of these five presuppositions, they cannot be included in heterodox economics.

Tony Lawson argues that realism is *the* key methodological distinguishing feature of heterodox economics, claiming that the other distinguishing factors derive from this one. While realism can be interpreted in many different ways, there is an agreement that the analysis must start with first approximations of the real world, not an idealistic one. This is in opposition to state-of-the-art orthodox models, the purpose of which is not even to represent reality.

Closely related to realism and instrumentalism is the kind of rationality that is assumed in our economic models. The only kind of rationality admissible to mainstream economists is constrained optimization with model-consistent expectations. Heterodox economics emphasize instead that economic agents in a world of fundamental uncertainty can at best satisfice some targets, adjusting to unfulfilled expectations by revising beliefs and norms.

The third pair of presuppositions concerns methodological individualism or atomicism versus holism or organicism. Orthodox economists believe that all analyses must start from the individual. By contrast, heterodox authors take a more holistic approach. They pay attention to the possibility of macroeconomic paradoxes, or fallacies of composition, such as Keynes's paradox of thrift.

About the fourth pair of presuppositions, readers should be reminded that the standard definition of orthodox economics is the study of scarcity. By contrast, heterodox economists are mostly concerned with unused productive capacity, the causes and the consequences of growth, and the ability of the economic system to create a surplus.

The fifth and last of our presuppositions is tied to the role of markets relative to that of the state. Mainstream economists exhibit confidence in the ability of unregulated markets to deliver stability and full employment. By contrast, heterodox economists are distrustful of unfettered markets. They suspect their inability to self-regulate, their tendency for destabilizing paths, and their squandering of resources. They believe that markets must be tamed. They argue that economic systems are not self-adjusting.

Some methodologists have an alternative view of what distinguishes orthodox and heterodox economics. They complain about the compulsive inclination of orthodox economists for formalistic modeling. They also argue that heterodoxy entails a belief in open systems, while orthodox economics needs to deal with closed systems.

Marc Lavoie

See also Austrian Economics; Critical Realism in Economics; Econophysics; Feminist Economics; Holism, in the Social Sciences; Individualism, Methodological; Institutional Economics; Marxist Economics

Further Readings

Lavoie, M. (2006). Do heterodox theories have anything in common? A post-Keynesian point of view. *Intervention: Journal of Economics, 3*(1), 87–112.

Lawson, T. (2006). The nature of heterodox economics. *Cambridge Journal of Economics, 30*(4), 483–505.

Lee, F. (2009). *A history of heterodox economics: Challenging the mainstream in the twentieth century.* London, England: Routledge.

O'Hara, P. A. (Ed.). (1999). *Encyclopedia of political economy.* London, England: Routledge.

O'Hara, P. A. (2007). Principles of institutional-evolutionary political economy: Converging themes from the school of heterodoxy. *Journal of Economic Issues, 41*(1), 1–42.

Historicism

Historicism refers to an array of approaches within philosophy and the social sciences that insists that history is central to any understanding of human affairs. It emerged as a response to modern and Enlightenment European philosophy's sometimes excessive enthusiasm for scientific method and causal explanation. In a variety of sometimes inconsistent ways, historicists maintain that human life and society must be understood historically. Their emphasis on the uniqueness of individual lives (rather than universal laws of human nature) in culturally varying contexts is often associated with both normative and theoretical relativism.

This entry traces the different meanings the notion of historicism took in the history of philosophy and the social sciences.

An important 18th-century source of historicism, Giambattista Vico defended the humanities and rhetoric against the exclusive claim of the natural sciences to valid knowledge, which was supported by philosophers from Francis Bacon and Descartes to David Hume. Vico maintains that what we know best is what we have made or created ourselves. There is a connection between what is "true" (*verum*) and what is "made" (*factum*), a connection reflected in our word *fact*. We are able to understand human actions, expressions, and artifacts with an immediate insight impossible with *nonhuman* entities. At the same time, our immediate insight into all things human does not guarantee objective knowledge.

As Johann Gottfried Herder made clear, human beings can only be understood through, in part because they are *constituted by,* their culturally and historically variable expressions. By implication, we must recognize as illusory that eternal and universal "human nature" so often invoked by Enlightenment thinkers as the object of social-scientific explanation. The languages and artifacts of different peoples, studied by disciplines such as philology, art history and literature, anthropology, and archaeology, always reflect their unique and irreducibly historical temperament or spirit (*Volksgeist*).

The realization that the truth of humanity is not to be found in an unchanging biology or psychology inspired some philosophers to seek that truth in the ultimate meaning or goal (*telos*) of history. Herder and Immanuel Kant both wrote ambitiously synoptic and speculative histories in that spirit. It was Georg Wilhelm Friedrich Hegel, however, who most influentially fashioned a wealth of historical insights and observations into a systematic philosophy of history. At the heart of Hegel's system is his view of history as a complex *dialectic*. Like the philosophical discourses (*dialektikê*) portrayed in Plato's Socratic dialogues, history advances through a series of mutually contending and superseding stages. He shows how different religions, works of art, political institutions, moral values, and metaphysical systems can be understood as stages in a complex dialectical process. Hegel's philosophy aspires to be the ultimate or "Absolute" truth gradually revealed through history.

Hegel's philosophical system is, in Hegel's own terms, "transcended" or "sublated" in the *materialist* dialectic of Karl Marx and Friedrich Engels. Where Hegel's conservative followers regarded the dialectic as substantially complete, Marx and Engels set themselves the radical task of bringing about the overthrow of present society in order to fulfill the *future* goal of history. At the same time, Hegel's intellectual or *idealist* dialectic of cultures and worldviews is transposed into the *materialist* register of modes of production and revolutionary class struggle.

Hegelian idealism and Marxism are two principal targets of Karl Popper's polemical engagement with historicism, which is deployed against all those theories and ideologies claiming knowledge of the essential meaning and future goal of history. Such approaches imply both historical determinism and methodological holism or collectivism. They are both politically and morally dangerous, because they treat individuals as unwitting dupes who may be sacrificed in order to guarantee humanity's historical mission. Historical determinism is intellectually incoherent because it presumes to anticipate a future society that depends, at least in part, on unforeseeable advances in human knowledge. Designed to combat now largely discredited totalitarian ideologies such as National Socialism and Soviet communism, Popper's critique can still be applied to more persistent Enlightenment notions of the inevitability of progress or "improvement." Another contemporary critic of the, mainly political, pernicious effects of historicism has been Leo Strauss.

Hegel's philosophy—freed of unsustainable assumptions of the Absolute—encouraged a wide

range of historicist approaches in the humanities and social sciences. Critical philosophers of history sought rigorous standards for the documentation and validation of historical knowledge. The 19th-century German philosopher Wilhelm Dilthey formulated rigorous criteria of "understanding" (*verstehen*) designed for all those *hermeneutic* disciplines primarily concerned with interpretation rather than causal explanation. The German historian Leopold von Ranke (late 18th to early 19th century) pioneered modern methods of historical scholarship in order to ensure an understanding of the past undistorted by present concerns and assumptions. In a parallel way, Marxism encouraged studies of social and economic history. At the same time, critical Marxists like Antonio Gramsci and members of the Frankfurt School set out to preserve the historicist insights of Marxism without succumbing to historical determinism (or "historicism" in Popper's sense).

In other disciplines, Franz Boas's anthropological historicism emphasized the uniqueness and incommensurability of different cultures. Legal historicists sought to understand law as neither the "natural" product of reason nor the arbitrary prescription of rulers, but rather as the specific historical expression of particular peoples.

Sometimes historicists claim that all knowledge is historical in similar ways. The Italian Neo-Hegelian philosopher Benedetto Croce and the British idealist R. G. Collingwood extended the historicist approach to philosophy and metaphysics. In the 20th century, Martin Heidegger and his student Hans-Georg Gadamer posited an essentially hermeneutic and so historical foundation for all human life and experience. Philosophical historicism of this kind tends toward moral, cognitive, and cultural relativism. In a similar vein, contemporary poststructuralists and postmodernists are typically skeptical, not only of any speculative philosophy or "grand metanarrative" of history but of all claims to universal truth. Michel Foucault's critical genealogy of history is probably the most influential of such approaches. Finally, "a new historicism" has appeared in literary theory.

David West

See also Determinism; Herder's Philosophy of History; *Naturwissenschaften* Versus *Geisteswissenschaften;* Philosophy of History; Popper's Philosophy of Science; Relativism and Social Science Straussian Critique of Social Science; Vico's *Scienza Nuova*

Further Readings

Dilthey, W. (2002). *The formation of the historical world in the human sciences*. Princeton, NJ: Princeton University Press.

Hegel, G. W. F. (1975). *Lectures on the philosophy of world history: Introduction* (H. B. Nisbet, Trans.). New York, NY: Cambridge University Press.

Marx, K., & Engels, F. (1975). *The communist manifesto*. New York, NY: Seabury.

Popper, K. R. (1961). *The poverty of historicism* (2nd ed.). London, England: Routledge.

Von Ranke, L. (2011). *The theory and practice of history.* London. England: Routledge.

Hobbes's Philosophical Method: Nature–Man–Society

This entry presents a brief introduction to the relationship between Thomas Hobbes's method and his investigations into nature, man, and society. When Hobbes (1588–1679) first conceived his political project, he thought to proceed from an investigation of nature to man to political society. However, as political tensions rose in his own society and addressing political questions became more urgent, he discovered that his political doctrine did not stand in need of any prior inquiry into larger nature or specifically human nature.

Hobbes immodestly wrote that the science of politics was no older than his own book *De Cive* (1642, in Latin). Earlier writers on politics had failed to discover the necessary properties of stable states, the rights of sovereigns, and the duties of subjects, because they lacked proper scientific method. Extrapolation from past experience may, if we have had a lot of experience and are very good at extrapolating from it, yield prudence; but this lacks the certainty of that sapience gained only by correct reasoning from proper principles. Experiences may be incomplete or wrongly interpreted; even sense perceptions may lead to false judgments, as when we perceive a straight stick as bent when partially submerged in clear water.

The proper method for discovering truth is *definitional.* We reason from premises that are analytically true, that is, true by virtue of the meanings of their component terms. The meanings of those terms must themselves enjoy a fixed, settled signification, and we are to proceed syllogistically using settled rules of inference. The conclusions we reach thus enjoy the status of reliable, noncontingent truths.

Of course the worry one may have about this method is of how to settle the appropriate component definitions. Conclusions will be warranted by their premises but, we might say, "garbage in, garbage out." Just any old arbitrary definitions will not track the reality of human concerns and experience or offer us a practically useful science of politics. Furthermore, many of the premises that must go into any realistic science of politics will have to include claims about what interests humans have, how they are motivated, how they interact in groups, and so on, which will not be true by definition but rather will be warranted by experience. It looks as if despite his stated definitional method, Hobbes will be forced to depend on some empirical observations if his political science is to have any interest for or claim on humans as we are.

One can address this apparent problem by reflecting on Hobbes's *reason* for aspiring to employ a definitional method. He wants his readers to have confidence in his conclusions and the system constructed on the basis of those conclusions. A definitional method, of the sort he perceived in Euclid's proof of the Pythagorean theorem—which Hobbes declared a totally convincing proof of an astonishing conclusion—was what he wanted for his own political argument. He wanted the component conclusions of his political philosophy not to be doubted, and not contestable, at least by anyone prepared to consider them carefully and in good faith. Hobbes (1994) wrote that

> to reduce this doctrine to the rules and infallibility of reason, there is no way, but, first put such principles down for a foundation, as passion, not mistrusting, may not seek to displace; and afterwards to build thereon the truth of cases . . . till the whole have been inexpugnable. (Epistle Dedicatory, p. 19)

Because Hobbes was engaged in the practical project of showing his countrymen that they have good and sufficient reasons to submit to the authority of their existing government, it makes sense to distinguish between those improperly synthetic premises that are harmful to his project and those that are not.

Presumably with this in mind, Hobbes permits himself some empirical premises, but only those that would not be doubted by anyone. One may think of such premises as *indubitable introspectables,* premises that "passion not mistrusting, will not seek to displace." Such premises will be premises that each person can confirm by introspection to be true in her own case, and so not doubt.

Such premises will not include universal generalizations about what everyone desires or what everyone thinks, because these cannot be known by individual introspection. Hobbes (1996) insists, first, that our science is too primitive to give us knowledge of human nature.

> For it is supposed that in this natural kingdome of God, there is no other way to know anything, but by natural reason; that is, from the principles of natural science; which are so farre from teaching us anything of God's nature, *as they cannot teach us our own nature,* nor the nature of the smallest creature living. (*Leviathan,* chap. XXXI, p. 33)

Hobbes's second reason for thinking that no proper proof of political principles will include empirical generalizations about what humans desire or think is principled. Science may tell us something of the *mechanisms* by which we operate—explaining what it is to want something or to believe something—but it cannot tell us the object, that is, what everyone wants or believes, because "for these [the objects of the passions] the constitution individual, and particular education do so vary, and they are so easie to be kept from our knowledge [by lying] that the characters of mans heart . . . are legible only to him that searcheth hearts" (Hobbes, 1996, *Leviathan,* Introduction)—that is, to God.

It has sometimes been thought that Hobbes assumed a very simple human psychology in which the aversion to temporal bodily death is the dominant, overriding desire of every sane human being; but Hobbes himself worried about the disruptions to social order caused by people's giving precedence to their desires to do their perceived religious duty, secure their salvation, or avenge an injustice. It is precisely because humans do not care most about securing the present life that they will engage in the seriously dangerous activity of waging war.

Hobbes may have been among the first, and arguably the most significant, of political philosophers to assert the independence of political theory. Rather than treating it as an application of the ethics for natural man, Hobbes treated it as the science of an *artificial entity,* the Commonwealth. As such, it stood in need of no prior scientific investigation into the natural world or into specifically human nature. Considerations of extrapolitical social dynamics do certainly play a role in his arguments as to what properties a political union, or commonwealth, must have in order to function well and remain stable over time. But these serve as background assumptions helping to specify the best definition of a commonwealth, from which Hobbes proceeds to derive his specific political conclusions.

Sharon A. Lloyd

See also Deduction; Individualism, Methodological; Introspection (Philosophical Psychology); Normativity; Rationality and Social Explanation; Scientific Method

Further Readings

Deigh, J. (1996). Reason and ethics in Hobbes's *Leviathan*. *Journal of the History of Philosophy, 34,* 33–60.

Gert, B. (2010). *Hobbes*. Cambridge, England: Polity Press.

Hobbes, T. (1994). Epistle dedicatory. In J. C. A. Gaskin (Ed. & Trans.), *Elements of law natural and politic* (pp. 19–20). Oxford, England: Oxford University Press. (Original work published 1640)

Hobbes, T. (1996). *Leviathan*. Cambridge, England: Cambridge University Press.

Hobbes, T. (1998). *On the citizen* (R. Tuck & M. Silverthorne, Ed. & Trans.). Cambridge, England: Cambridge University Press. (Original work published 1642 in Latin as *De Cive*)

Johnston, D. (1986). *The rhetoric of "Leviathan": Thomas Hobbes and the politics of cultural transformation*. Princeton, NJ: Princeton University Press.

Lloyd, S. A. (2009). *Morality in the philosophy of Thomas Hobbes: Cases in the law of nature*. Cambridge, England: Cambridge University Press.

Lloyd, S. A. (Ed.). (2012). *Hobbes today: Insights from the 21st century*. Cambridge, England: Cambridge University Press.

Martinich, A. P. (2005). *Hobbes*. New York, NY: Routledge.

Pettit, P. (2008). *Made with words: Hobbes on language, mind and politics*. Princeton, NJ: Princeton University Press.

Holism, in the Philosophy of Language

Holism covers a wide variety of theses, according to each of which the whole of a theory or a language or a system has some property that its parts lack and, further, this property cannot be defined or otherwise characterized by the properties of these parts. In epistemology, holism is the view that whole theories are the smallest units of confirmation. Single hypotheses yield observational predictions only with the aid of a body of background theory. This means that a failed prediction does not conclusively refute the hypothesis from which it is derived. This sort of holism has been defended by many 20th-century thinkers, including Pierre Duhem, W. V. O. Quine (cf. his phrase "the web of belief"), and Rudolf Carnap. The transition from epistemic holism to meaning holism was pushed by empiricist philosophers who wished to explain the meaning of symbols in terms of their verification conditions. Meaning holism (with variants such as "semantic holism" and "linguistic holism"), then, became the view that more inclusive bodies of theory serve as the basic units of meaning, and not the intuitive position on words as the fundamental bearers of meaning.

Meaning Holism

Historically, philosophers have been divided over whether meaning properties of expressions, such as a linguistic item expressing a concept or a proposition, or having a referent, or truth conditions, are holistic or atomic.

Those in the atomist tradition proceed from British empiricism to American pragmatism, the *locus classicus* being the word, as in the Vienna Circle; its contemporary proponents include most model theorists (Richard Montague, John Perry, and Jon Barwise), behaviorists (e.g., the psychologists B. F. Skinner and John B. Watson), and those philosophers whose work concerns so-called informational semantics (e.g., Fred Dretske, Jerry Fodor, and Ruth Millikan). People in these traditions think that the semantic properties of a symbol are determined by its relations to things in the nonlinguistic world.

The holist tradition, on the other hand, proceeds in linguistics from the structuralist tradition

and in philosophy from Ludwig Wittgenstein, W. V. O. Quine, Donald Davidson, Daniel Dennett, Ned Block, Hilary Putnam, Richard Rorty, Wilfred Sellars, and Robert Brandom; and it further includes almost everyone in the fields of artificial intelligence (AI) and cognitive science. People in this tradition hold that the semantic properties of a symbol are determined by its role in a language. So construed, meaning holism is a metaphysical thesis about the nature of representation in which the meaning of a symbol is relative to the entire system of representations containing it. Thus, a linguistic expression can have a meaning only in the context of a language; analogously, a hypothesis can have significance only in the context of a theory; and similarly, a concept can have intentionality only in the context of the belief system.

Meaning holism has profoundly influenced virtually every aspect of contemporary theorizing about language and mind, not only in the philosophy of language and the philosophy of mind but also in anthropology, linguistics, literary theory, AI, psychology, and cognitive science.

According to W. V. O. Quine, there can be no nonquestion, begging way of distinguishing (analytic) statements that are true in virtue of meaning alone from (synthetic) statements whose truth depends on facts about the world beyond meaning. This thesis serves as a premise in practically all arguments for holism about meaning. If the meaning of an expression is determined partly by its role in a language and if the analytic/synthetic distinction is infirm, then there can be no principled distinction between those aspects of a word's linguistic role that are relevant to determining its meaning and those that are not. The invited inference, obviously, is that the meaning of a word is a function of its whole linguistic role in the language. As Quine (1951) famously put it, "The unit of empirical significance is the whole of science" (p. 42).

Because holism, as we are construing it, is a metaphysical thesis, it is not a semantic (empirical) one, and so two theories might agree about the semantic facts but disagree about holism. For example, nothing in the logician Alfred Tarski's writings determines whether the semantic facts expressed by the theorems of an absolute truth theory are holistically determined or not. Yet Davidson, an ardent semantic holist, argued that the correct form for a semantic theory for a natural language *L* is an absolute truth theory for *L*. Semantic theories, like other sorts of theories, need not wear their metaphysical commitments on their sleeves.

Meaning holism has startling implications: Suppose that you think Barack Obama owned a dog and I think he did not. Because our beliefs differ, there are inferences about dogs you are prepared to accept and I am not. But holism says that what "dog" means in your mouth depends on the totality of your beliefs about dogs, including your beliefs concerning who owns them. It would thus seem to follow that you and I must mean different things by our shared word *dog*. This line of argument leads, more or less directly, to surprising consequences—for instance, that natural languages are not, in general, intertranslatable (W. V. O. Quine, Ferdinand de Saussure); that there may be no fact of the matter about the meanings of texts (Hilary Putnam, Jacques Derrida); and that scientific theories that differ in their basic postulates are "empirically incommensurable" (Paul Feyerabend, Thomas Kuhn). Moreover, *meaning holism* seems to imply that there cannot be a science of mental phenomena. For if meaning holism is true, then no two people can have the same beliefs or desires. But then no two people can fall under the same psychological law; indeed, no two time-slices of the same person can fall under the same psychological law.

Reactions to Holism

Philosophers react to holism and its implications in a variety of ways. A number of them just bite the bullet: Strictly speaking, they say, there are no such things as mental states, no such things as determinate translations or shared psychologies. Variations on this skepticism can be found in W. V. O. Quine, Donald Davidson, Daniel Dennett, Hilary Putnam, Stephen Stich, and Paul and Patricia Churchland, among others. These philosophers are sometimes called *semantic nihilists*.

Semantic nihilism is, of course, the most radical response to the consequences of holism. It is the view that, strictly speaking, there are no semantic properties. Strictly speaking, there are no mental states; words lack meanings. At least for scientific purposes (and perhaps for other purposes as well), we must abandon the notion that most people are moral or rational agents and that they act on their beliefs and desires.

Other philosophers have sought to avoid the dreaded consequences of holism by developing a

sort of mitigated holism according to which a notion of similarity of meaning (of mental or linguistic content) somehow replaces the notion of meaning identity that appeals to the analytic/synthetic distinction. This approach is also popular among cognitive scientists (Hartry Field, Gilbert Harman, and Paul Smolensky, among others). And still others suggest that perhaps an analytic/synthetic distinction can be preserved from Quine's attack by developing a "graded" notion of analyticity (Ned Block, Michael Devitt) or by arguing that there must be a viable analytic/synthetic distinction because the consequences of there not being one are simply too awful to contemplate (Michael Dummett). These philosophers are sometimes called *semantic molecularists*.

Semantic molecularism, like semantic holism, holds that the meaning of a representation in a language *L,* is determined by its relationship to the meanings of other expressions in *L* but, unlike holism, not by its relationships to every other expression in *L*. Molecularists, instead, are committed to the view, contra Quine, that for any expression *e* in a language *L,* there is an in-principle way of distinguishing between those representations in *L* the meanings of which determine the meaning of *e* and those representations in *L* the meanings of which do not determine the meaning of *e*. Traditionally, this in-principle delimitation is supposed by an analytic/synthetic distinction. Those representations in *L* that are meaning constituting of *e* are analytically connected to *e,* and those that are not meaning constituting are synthetically connected to *e*. Meaning molecularism seems to be the most common position among philosophers who reject meaning holism.

A different sort of reaction to holism is *atomism*. Meaning atomists hold that the meaning of any representation (linguistic, mental, or otherwise) is not determined by the meaning of any other representation. Historically, for example, Anglo-American philosophers in the 18th and 19th centuries thought that an idea of an *X* was *about* *X*s in virtue of this idea's physically resembling *X*s in the world. These sorts of resemblance theories, of course, are no longer fashionable, but a number of contemporary meaning atomists continue to believe that the basic semantic relation is between a concept and the things to which it applies, not an inferential relationship among the concepts themselves.

Ernest Lepore

See also Analytic/Synthetic Distinction; Concepts; Duhem-Quine Thesis and the Social Sciences; Holism, in the Social Sciences; Inferentialism; Language, Philosophy of; Semantics and Pragmatics; Teleosemantics

Further Readings

Block, N. (1988). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy, 12,* 615–678.

Brandom, R. (1994). *Making it explicit*. Cambridge, MA: Harvard University Press.

Carnap, R. (1937). *The logical syntax of language*. London, England: Routledge & Kegan Paul.

Davidson, D. (1984). Belief and the basis of meaning. In *Inquiries into truth and interpretation* (pp. 141–154). Oxford, England: Clarendon Press.

Duhem, P. (1962). *The aim and structure of physical theory*. New York, NY: Atheneum.

Field, H. (1977). Logic, meaning and conceptual role. *Journal of Philosophy, 74,* 379–409.

Fodor, J. (1987). *Psychosemantics*. Boston, MA: MIT Press.

Fodor, J., & Lepore, E. (1992). *Holism: A shopper's guide*. Oxford, England: Blackwell.

Quine, W. V. O. (1951). *Pursuit of truth* (chap. 1). Cambridge, MA: Harvard University Press.

Quine, W. V. O. (1980). Two dogmas of empiricism. In *From a logical point of view* (2nd ed.). Cambridge, MA: Harvard University Press. (Reprinted from *Philosophical Review, 60,* 20–43, 1951)

Holism, in the Social Sciences

Within the debate on holism in the philosophy of social sciences, it is common to distinguish between two issues. One is the ontological question of whether social phenomena exist, or exist in their own right. Ontological holists maintain that they do, whereas ontological individualists deny this. The other is the methodological question of the proper focus of explanations. Here, methodological holists hold that the social sciences should offer holist explanations, that is, explanations in terms of social phenomena. By contrast, methodological individualists insist that the social sciences should provide only individualist explanations, that is, explanations exclusively in terms of individuals, their actions, beliefs, desires, and so on.

The debate on holism has mainly concentrated on the methodological issue. Accordingly, this will

be the main focus of the present entry. The first section contains a brief presentation of ontological holism. Next follows a more thorough outline of methodological holism. Finally, various arguments in support of methodological holism are examined.

The Thesis of Ontological Holism

Ontological holism is sometimes characterized as the view that social phenomena exist. More often, however, it is phrased as the claim that social phenomena exist in their own right or as sui generis. Two aspects of this thesis deserve further comment.

To begin with, it is common to distinguish between different kinds of social phenomena, such as organizations (e.g., a school or a nation), the statistical properties of a group (e.g., its literacy rate or suicide rate), and the rules and norms within a group (e.g., the rule to drive to the right). The debate on ontological holism has mainly concentrated on social phenomena in the form of organizations.

Regarding the notions of existence and sui generis existence, there are various stances as to what exactly it takes for social phenomena to enjoy existence or sui generis existence. For instance, social phenomena may be said to exist in their own right (a) if they are composed of something more than constellations of individuals; (b) if they have causal powers that are independent of, and outstrip, the causal powers of individuals; or (c) if predicates such as "school" or "nation" that refer to social phenomena cannot be translated into statements about individuals. As a result of there being multiple criteria of existence and sui generis existence, ontological holists and ontological individualists may—and often do—have different things in mind when discussing whether social phenomena exist or exist sui generis.

Ontological holism is compatible with both a holist and an individualist position on the methodological issue. Thus, a commitment to ontological holism does not in itself entail a commitment to methodological holism, and vice versa. It is only when a given criterion of existence or sui generis existence is combined with suitable assumptions that a holist stance on ontology implies a holist stance on methodology, and vice versa. Within the individualism/holism debate, this insight did not receive much attention until the 1950s. At that point, the distinction between the ontological and methodological dimension of the debate began to gain currency. Equipped with this distinction, many methodological holists began to stress that they were not committed to, and did not endorse, the thesis of ontological holism. They distanced themselves from the view that social phenomena are composed of something more than constellations of individuals, just as they rejected the contention that social phenomena have causal powers that are independent of, and outstrip, the causal powers of individuals. Today, the majority of participants in the debate, methodological holists included, consider themselves to be ontological individualists. Typically, they take this to mean that social phenomena are nothing but mere constellations of individuals.

The Thesis of Methodological Holism

Methodological holism is the view that holist explanations, that is, explanations in terms of social phenomena, should be advanced within the social sciences. As it stands, this thesis raises three issues in need of clarification.

1. *First, when does an explanation qualify as being holist, that is, as being in terms of social phenomena?* There are various stances on this matter. For instance, one is that an explanation is holist if the explanans, that is, what does the explaining, refers to social phenomena. Another is that an explanation is holist if the explanans contains social predicates or social descriptions referring to corresponding social phenomena. In turn, both these proposals may be further—and differently—spelled out depending on what counts as social phenomena and what as social predicates, respectively. This point may be illustrated in relation to the suggestion that an explanation is holist if the explanans contains social predicates used as a reference. Here, it is generally agreed that social predicates are exemplified by descriptions such as "school," "government," "a decrease in unemployment," "a rise in crime," "the rule that . . . ," and "the norm that. . . ." At the same time, there are other types of predicates where it is a matter of dispute whether they should be classified as social descriptions. Most notably, some insist that social predicates should be used broadly to include descriptions of relations between individuals and role predicates such as "bank clerk" and "policeman." On the contrary, others argue that these predicates should be regarded as descriptions that may figure in individualist

explanations. As these considerations show, it is possible to distinguish between broader or narrower conceptions of what to count as a holist explanation. The broader conceptions draw the line between holist and individualist explanations such that more explanations come out as holist explanations. The narrower ones regard fewer explanations as holist by allowing more explanations to qualify as individualist. Very often, methodological holists subscribe to broader conceptions, whereas methodological individualists favor narrower ones. As a result, they tend to talk past each other: Each offers arguments presupposing a conception of holist explanations that is unacceptable to their opponent. This being the case, their arguments fail to impress the other.

2. *Second, what is meant by an explanation?* Again, there are different answers to this question since there are various notions of what to count as an explanation. Consequently, methodological holism may be variously spelled out depending on the notion of explanation adopted. Much of the earlier debate is conducted in terms of the covering-law model of explanation. According to this model, a scientific explanation takes the form of a deductive or inductive argument that shows why the phenomenon in need of explanation was to be expected. For quite a while now, the covering-law model has been under heavy attack, to say the least. Not surprisingly, therefore, more recent specifications of methodological holism typically rely on alternative notions of explanation. For instance, they subscribe to the causal-information view of explanation, which states that an explanation provides information about the causal process leading to the phenomenon in need of explanation. Or they embrace the erotetic model of explanation, which states, roughly speaking, that an explanation is an answer to a why-question. It is important to note here that irrespective of the notion of explanation adopted, methodological holists regard holist explanations as finished and complete explanations. Thus, they deny that holist explanations are only tolerable as unfinished explanatory stopping points on the way to individualist explanations considered as finished explanations. Likewise, they reject the stance that holist explanations are incomplete unless they are supplemented by accounts in terms of individuals. To be a methodological holist is to hold that holist explanations are just fine as they stand.

In addition to diverse *notions* of explanation, it is possible to distinguish between various *kinds* of explanation. Two examples of different kinds of explanation are functionalist and straightforward causal explanations. Historically speaking, methodological holism has often been associated with the advancement of functionalist explanations. Applied by methodological holists, these assert that the continued existence of some social phenomenon is explained by its function or effect. For instance, it may be suggested that the state continues to exist because it furthers the interests of the ruling class. Nowadays, functionalist explanations are generally regarded as having a very limited applicability to social phenomena. Currently, the kind of explanation that is mostly used is probably the straightforward causal explanation. It is exemplified by the claim that the rise in unemployment caused the rise in crime and that the government's decision to lower the taxes led to an increase in the consumption of luxury goods. Needless to say, it is possible to distinguish further kinds of holist explanations, just as more fine-grained or completely different classifications of kinds of holist explanations may be adopted. The methodological holist who relies on a given *notion* of explanation may, to a varying extent, make use of different *kinds* of explanation.

3. *Third, to what extent should holist explanations be advanced within the social sciences?* As formulated above, the thesis of methodological holism leaves this question open. One option is to further specify that holist explanations *alone* should be offered within the social sciences. Individualist explanations should not be put forward. Among methodological holists, this is the minority view. Most of them adopt the more moderate stance that *on some occasions* holist explanations are in order. On other occasions, individualist explanations may well be provided.

As these considerations bring out, it is possible to distinguish between numerous versions of methodological holism depending on how the thesis is spelled out. In the 1950s, some versions of the thesis were strongly opposed by methodological individualism. As a result, methodological holism came to be seen as a position on the defensive. Methodological individualism was the dominant position. Today, this is no longer obviously the case. Or at the very least, it may be registered that a number of influential

theorists are currently defending their favorite version of the holist stance on explanation.

Arguments in Support of the Indispensability of Holist Explanations

Different arguments have been offered in support of methodological holism. Here follows an examination of four main arguments. In different ways, they purport to show that holist explanations are indispensable.

Arguments From Causal Powers

One type of argument appeals to the causal powers of social phenomena. It comes in different versions. One version proceeds like this: Social phenomena exist sui generis in the sense of having causal powers that are somehow independent of, and outstrip, the causal powers of individuals. This means that when social phenomena exercise their independent and outstripping causal powers, their effects may not be explained merely by appealing to individuals with their limited causal powers. Instead, it is necessary to offer explanations that refer to the social phenomena that brought about the effects. In short, only holist explanations will do. This being the case, on pain of leaving these effects unexplained, the social-scientific use of holist explanations is indispensable.

An argument along these lines is sometimes implicit or explicit in the writings of theorists working within a Hegelian, Marxist, or functionalist tradition. Today, the argument does not enjoy much, if any, support. It is simply too hard to make sense of, and render plausible, the contention that social phenomena have causal powers that are independent of, and outstrip, the causal powers of individuals.

Another version of the argument from causal powers avoids this problem. Here, it is acknowledged that the causal powers of social phenomena depend on, and do not outstrip, the causal powers of single individuals. Still, it is held that the causal powers of social phenomena are distinct from the causal powers of single individuals. Consequently, when social phenomena exercise their distinct causal powers, explanations of their effects must point to these social phenomena. The explanations offered must be holist. It will not do to use individualist explanations since single individuals do not bring about the effects in question. Therefore, in order not to leave the effects unexplained, holist explanations must be advanced within the social sciences.

This version of the argument is currently being defended by some of the theorists working within the tradition of critical realism. Among other things, the argument has been criticized for not presenting convincing reasons to the effect that the causal powers of social phenomena are indeed distinct from the causal powers of individuals.

Arguments From the Impossibility of Translation

Another type of arguments focuses on whether holist explanations may be translated into individualist explanations. More specifically, it is first pointed out that holist explanations contain social predicates. Then, it is maintained that the meaning of social predicates cannot be captured solely in terms of descriptions of individuals. Or, as it is also put, social predicates are not exhaustively analyzable, or reductively definable, in terms of descriptions of individuals only. This being the case, it is impossible to translate holist explanations into individualist explanations. Finally, it is then concluded that since holist explanations may not in this manner be replaced by individualist explanations, holist explanations are indispensable.

But why hold that this translation is impossible? The most famous and widely discussed argument to this effect was advanced in the 1950s. It states that the definition of social predicates will inevitably contain other social predicates. For instance, the definition of "bank teller," here considered as a social predicate, must mention a bank. And the definition of "bank" must, in turn, contain notions such as "legal tender" and "contract." And these social predicates, in turn, may only be defined in ways that involve yet other social descriptions, and so on. In this fashion, it is impossible to capture the meaning of social predicates exclusively in terms of descriptions of individuals.

Today, there are few, if any, proponents of this argument. Instead, other reasons are given to show that holist explanations cannot be translated into individualist ones. This point, though, does not receive much attention in the current debate. It is generally held that in order to show that holist explanations are indispensable it does not suffice to demonstrate that they cannot be translated into individualist ones.

Arguments About the Impossibility of Intertheoretic Reduction

A third type of arguments purports to demonstrate the impossibility of intertheoretic reduction. Arguments of this sort take it that holist explanations always draw on holist theories, whereas individualist explanations always involve individualist theories. Against this background, it is then maintained that holist theories are oftentimes irreducible to individualist theories. By implication, holist explanations that make use of irreducible holist theories cannot be substituted by individualist explanations, which involve individualist theories. Thus, on pain of leaving various phenomena unexplained, the use of these holist explanations is indispensable within the social sciences.

The notion of intertheoretic reduction may be specified by applying a standard model of reduction to holist theories. A holist theory contains social predicates. It must first be shown that its social predicates may be linked, on a one-to-one basis, with descriptions of individuals via bridge laws. The bridge laws express that the linked predicates are coextensive, that is, have the same reference, in a lawlike manner. Then, the holist theory must be deduced from, and in this sense explained by, an individualist theory plus the bridge laws. When these conditions are met, the holist theory has been reduced to an individualist theory.

One of the arguments that methodological holists have offered against the possibility of intertheoretic reduction is the argument from multiple realization. Applied by methodological holists, the argument states that many, if not most, types of social phenomena may be realized by multiple types of constellations of individuals. When this is the case, the corresponding social predicates may not be linked up with single descriptions of individuals via bridge laws. For instance, consider bureaucracies: Individuals may be organized in multiple ways while still realizing or constituting a bureaucracy. Consequently, the social predicate "bureaucracy" must be linked not with a single description but with multiple descriptions of how individuals may realize this kind of social phenomenon. Whenever this happens, the first condition of intertheoretic reduction is not met. Intertheoretic reduction fails.

Methodological holists began to advance the argument from multiple realization in the early 1980s. Presently, it is regarded as *the* argument against reduction. However, the argument has also been disputed on various accounts. Moreover, the model of reduction and its applicability to the debate about methodological holism has been challenged.

The Argument From Explanatory Interests

The last type of argument to be considered here has the following structure: At first, it is made clear that a certain notion of explanation is adopted. By appeal to this notion, it is then argued that a holist explanation may serve interests that are not satisfied by an individualist explanation of the same particular event or state. This being the case, it is concluded that since the social sciences should serve these interests, holist explanations are indispensable.

Arguments of this type have mainly been advanced since the early 1980s while relying on either the causal-information or the erotetic view of explanation. According to the former, an explanation provides information about the causal process leading to the event or state in need of explanation. Equipped with this notion of explanation, consider an example of a holist and an individualist explanation of the same particular event. The methodological holist might explain the rise in crime in Copenhagen last year by pointing to a rise in unemployment. By contrast, the methodological individualist might say that the crime rate went up because certain individuals had lost their jobs and felt frustrated about having very little money and no job opportunities. The holist explanation conveys that given the rise in unemployment, a rise in crime was almost bound to occur. If those particular individuals had not lost their job and begun to commit crimes, then others would have. As such, it provides a different sort of information than the individualist explanation. The individualist explanation confines itself to pointing out that these particular individuals with a changed job situation and incentives were responsible for the rise in crime. In this fashion, the example illustrates how holist and individualist explanations satisfy different interests in the causal history of the same particular events or states.

Among other things, the argument has been criticized on the ground that it does not sufficiently bring out that the dissimilar causal information provided by holist and individualist explanations

may serve different theoretical *as well as* pragmatic interests. Also, it has been argued that the argument from explanatory interests is better made by relying on the erotetic notion of explanation, that is, the view that an explanation is an answer to a why-question.

Julie Zahle

See also Collective Agents; Durkheim's Philosophy of Social Science; Emergence and Social Collectivism; Explanation, Theories of; Group Mind; Holism, in the Philosophy of Language; Individualism, Methodological; Microfoundationalism; Plural Subjects; Reductionism in the Social Sciences; Social Facts; Social Ontology, Recent Theories of; Structural Functionalism, in Social Theory

Further Readings

Elder-Vass, D. (2007). For emergence: Refining Archer's account of social structure. *Journal for the Theory of Social Behaviour, 37*(1), 25–44.

Epstein, B. (2009). Ontological individualism reconsidered. *Synthese, 166*(1), 187–213.

Garfinkel, A. (1981). *Forms of explanation.* New Haven, CT: Yale University Press.

Jackson, F., & Pettit, P. (1992). In defense of explanatory ecumenism. *Economics and Philosophy, 8,* 1–21.

James, S. (1984). *The content of social explanation.* Cambridge, England: Cambridge University Press.

Kaldis, B. (1993). *Holism, language and persons: An essay on the ontology of the social world.* Aldershot, England: Ashgate.

Kincaid, H. (1996). The failure of individualism. In *Philosophical foundations of the social sciences* (chap. 5, pp. 142–190). New York, NY: Cambridge University Press.

Kincaid, H. (1997). *Individualism and the unity of science.* Lanham, MD: Rowman & Littlefield.

Macdonald, G., & Pettit, P. (1981). *Semantics and social science* (chap. 3, pp. 106–149). London, England: Routledge.

O'Neill, J. (Ed.). (1973). *Modes of individualism and collectivism.* London, England: Heinemann.

Pettit, P. (1993). *The common mind: An essay on psychology, society and politics.* Oxford, England: Oxford University Press.

Weber, E., & Van Bouwel, J. (2002). Can we dispense with structural explanations of social facts? *Economics and Philosophy, 18,* 259–275.

Zahle, J. (2003). The individualism-holism debate on intertheoretic reduction and the argument from multiple realization. *Philosophy of the Social Sciences, 33*(1), 77–99.

Zahle, J., & Collin, F. (Eds.). (forthcoming). *Rethinking the individualism/holism debate: Essays in the philosophy of social science.*

Homo Economicus

Homo economicus is a model of human behavior that explains social phenomena (including social interaction outside the market) in terms of economic behavior or what the received view construes as rational economic behavior: namely, as opportunity taking or (opportunistic) rational action. This basically Hobbesian model of man underlies most of the modern, rational choice approaches to explaining social phenomena. The choice-making behavior of homo economicus is characterized by future directedness, case-by-case motivation, and subjectivism.

Rational economic (wo)men strictly distinguish between what is and what is not a causal consequence of each of their acts taken separately. A rational person complies with a norm if and only if compliance is judged to have better consequences than noncompliance. Intrinsic motivation by the norm or rule is not operative. Whether it would be good or bad for a choice maker to act always in the same way or whether it would be good or bad if all would act in a specific way is irrelevant. Any regularity in choice-making emerges from the fact that the same external incentives operate on the opportunistic choice-making entity regularly.

Consistency Assumption

Whatever it is that the rational individual/homo economicus does, as long as choice making fulfills the axioms that guarantee "consistency"—it is not possible to make somebody pay to get A for B and have the same person pay to trade B for A—the behavior can be described "as if" utility maximizing, by assigning higher utility indicators to the (consistently) chosen alternatives than to the nonchosen. Egoistic as well as nonegoistic choice behavior can be represented "as if" maximizing. For, according to the classical model of "utility maximization,"

an alternative is preferred because it has higher utility. According to the modern model, utility is not among the reasons for choice. By choosing an alternative—for whatever (un)selfish motives—the preference is assumed to be revealed. Higher utility is assigned to the chosen alternative merely to indicate the preference. Homo economicus behaves as if she were maximizing a utility function, yet the higher or lower values of that function are not among the reasons for preferring.

Case-by-Case Motivation

The *core assumption* of case-by-case opportunistic—rather than rule-guided—choice making in view of the *causal future consequences of each act taken separately* goes beyond the *consistency assumption.* It has far-reaching implications. These can best be illustrated in terms of the theory of (repeated) games. As a specific example, take a classical one-off, two-by-two Prisoner's Dilemma (PD) game. In this game, moves are made separately and in isolation and therefore cannot causally influence each other. No player can make her own choice causally dependent on that of the coplayer. Each actor has a dominant choice alternative that leads to better results for her regardless of what the coplayer chooses. If both actors do what is better for them regardless—that is, act rationally—they end up in a (Pareto inferior) situation to which another situation exists in which both would fare better.

As long as each actor assumes that the choices of the other individual are causally independent from her own, nothing will alter the dominance characteristics of the actions in the basic PD game. However, if the PD is repeated, actions performed on a former round of play can causally influence actions on later rounds of play. It is often thought that through this mechanism the so-called Hobbesian problem of explaining social order in terms of case-by-case maximizing of future-directed behavior could be solved. In repeat interaction, it is claimed, homo economicus would rationally show reciprocity: If *other* cooperated on the last, *self* will do so on the next round of play, and if *other* defected on the last, *self* reciprocates on the next. Then, opportunistic behavior would seem as if guided by "norms" of reciprocity. However, though such "tit-for-tat" behavior has a basis in the retributive emotions, it is not for strictly rational homo economicus. Taking seriously the future directedness of rational choice making the past behavior can never matter as such (bygones are bygones).

In a finitely repeated game, on the last round the dominance properties of the normal game apply. There is no future beyond the last round. Therefore, no relevant causal influence could induce individuals to deviate from their dominant strategy choices on the last round. This infects the next to last round, and so on. Rational reciprocal cooperation would unravel among rational actors who would foresee this.

If the model of rational economic man is taken seriously and only causal influences on the future matter, then any past history of play must be irrelevant. This applies as well if the game is identically and indefinitely repeated in the future. After removing at most finitely, many initial rounds of play still infinitely many remain. *Independent of history,* the future remaining game—except for renumbering—is structurally identical. For rational economic (wo)men it follows that in view of the structurally identical future they should always make the same choices regardless of the past. Strategies such as "tit for tat," which make behavior contingent on behavior on previous rounds of play will not make sense, and the alleged rational choice solutions of the Hobbesian problem of social order fail.

Homo economicus acts upon a model of the situation according to what she judges right with respect to future consequences. This brings "teleology" into a causal world. Distinguishing between what is and what is not a causal consequence is itself a process subject to causal laws, however. Homo economicus will explore these and other laws that govern the process by which the past transforms itself into the expected future.

Hartmut Kliemt

See also Decision Theory; Equilibrium in Economics and Game Theory; Hobbes's Philosophical Method: Nature–Man–Society; Pareto Optimality; Rational Expectations; Rationality and Social Explanation; Sen's Paretian Liberal

Further Readings

Binmore, K. G. (2009). *Rational decisions.* Princeton, NJ: Princeton University Press.

Güth, W., & Kliemt, H. (2007). The rationality of rational fools. In F. Peter & H. B. Schmid (Eds.), *Rationality and commitment* (pp. 124–149). Oxford, England: Oxford University Press.

Hobbes, T. (1968). *Leviathan.* Harmondsworth, England: Penguin. (Original work published 1651)

Parsons, T. (1968). Utilitarianism: Sociological thought. In D. Sils & R. K. Merton (Eds.), *International encyclopedia of social sciences* (Vol. 16, pp. 229–236). New York, NY: Macmillan.

Pettit, P. (1995). The virtual reality of *Homo Economicus. The Monist, 78*(3), 308–329.

Human Cultural Niche Construction and the Social Sciences

Niche construction theory (NCT) is a novel theoretical perspective in evolutionary biology that has challenged the hitherto dominant, one-sided view of Neo-Darwinian evolution, with gene-driven evolution as the sole protagonist. NCT posits an alternative paradigm to human sociobiology or evolutionary psychology. It has important repercussions beyond biology as such, being also applicable to the human species and its cultural practices, showing that human cultural interventions and social processes can have feedback effects on the selection pressures on genes. NCT can, furthermore, have an impact on the human and social sciences, affecting the way social explanation works. Some go so far as to claim that "niche construction theory potentially integrates the biological and social aspects of the human sciences" (Kendal et al., 2011).

Niche construction is the process whereby organisms, through their activities and choices, modify their own and each other's niches. By transforming natural selection pressures, niche construction generates feedback in evolution, on a scale hitherto underestimated and in a manner that alters the evolutionary dynamic. Advocates of the niche construction perspective seek to develop a new approach to evolution—one that treats niche construction as a fundamental evolutionary process in its own right. This approach has become known as NCT.

Niche construction also plays a critical role in ecology, where it supports ecosystem engineering and eco-evolutionary feedbacks and partly regulates the flow of energy and nutrients through ecosystems. NCT is also starting to have an impact in a variety of other disciplines, including the human sciences, philosophy of biology, medicine, developmental biology, and conservation biology. In any one discipline, the effect of NCT may still yet be modest, but across them collectively, it is starting to become a powerful interdisciplinary movement.

NCT is also relevant to the philosophy of science because it challenges certain fundamental aspects of orthodox Neo-Darwinism.

The Reference Device Problem

The evolutionary biologist Richard Lewontin drew attention to a problem within evolutionary biology by summarizing standard evolutionary theory as follows:

$$\frac{dO}{dt} = f(O, E) \tag{1a}$$

$$\frac{dE}{dt} = g(E) \tag{1b}$$

Evolutionary change, dO/dt, is assumed to depend on both organisms' states, O, and environmental states, E (Equation 1a), but environmental change, dE/dt, is assumed to depend on environmental states only (Equation 1b). With many caveats and complications, organisms are not generally regarded as causing evolutionarily significant changes in their environments. While organismal change of environments is clearly recognized by biologists to occur, this has typically been assumed to operate on temporal and spatial scales that are irrelevant to evolutionary analysis. The exceptions are when the selective environment is provided by other organisms, as in the case of coevolution or sexual selection.

For Lewontin, a more accurate general description of how evolution actually works is as follows:

$$\frac{dO}{dt} = f(O, E) \tag{2a}$$

$$\frac{dE}{dt} = g(O, E) \tag{2b}$$

where environmental change also depends on the environment-modifying activities of organisms. The philosopher of biology P. Godfrey-Smith drew

attention to the same problem by describing standard evolutionary theory as "externalist," by which he meant that it seeks to explain the internal properties of organisms, their adaptations, exclusively in terms of external properties, that is natural selection pressures, in their external environments. The principal point that the conventional approach obscures is that organisms change some of the selection pressures in their environments.

The "reference device problem" is that the causal arrow in Equation 2 representing niche construction points in the "wrong" direction, from organisms to environments, and is therefore not compatible with the externalist assumption of standard evolutionary theory. As a result, it is difficult for evolutionary biologists to describe changes in selection caused by niche construction as evolutionarily causal. Instead, standard evolutionary theory is forced by its own explanatory reference device to "explain away" all observed instances of niche construction as phenotypic, or extended phenotypic, consequences of prior natural selection. Standard evolutionary theory can recognize niche construction as a consequence of evolution, but it cannot recognize it as a cause in its own right.

The Devaluation of Proximate Causes

Responding to structuralist critics, in 1984 Ernst Mayr wrote, "All of the directions, controls and constraints of the developmental machinery are laid down in the blueprint of the DNA genotype as instructions or potentialities" (p. 126). For Mayr, developmental processes cannot be regarded as independent causes of evolutionary events, since their characteristics, including their ability to control and constrain, are themselves fully explained by prior natural selection. If developmental processes direct evolutionary events, this is only the *proximate* manifestation of the *ultimate* cause of natural selection; conversely, those aspects of development that have not been shaped by selection play no evolutionary role.

Mayr was extremely influential in bringing this commonly made distinction between "proximate" and "ultimate" causes to prominence within biology. In an article published in 1961, Mayr argued that natural selection should be regarded as the *ultimate* cause of phenotypic characters, thereby effectively devaluing so-called proximate causes as explanatory tools within evolutionary biology. Since niche construction includes developmental processes, this stance also prevented evolutionary biologists from recognizing niche construction as an evolutionary process and thus hinders the integration of evolutionary and developmental biology.

Instead, niche construction is perceived to have no *independent* evolutionary significance because, to the extent that it is evolutionarily consequential, it is regarded as fully explained by a preceding cause, natural selection. Niche construction effects are treated as merely extended phenotypes, and extended phenotypes play the same role in evolutionary biology as ordinary phenotypes—namely, to affect the replication potential of the alleles contributing to those phenotypic effects. While this stance recognizes that modification of the selective environment does occur, it does not view environmental modification as a process with quasi-independent causal significance.

There are two major problems with this line of reasoning. First, not all evolutionarily consequential niche construction, or all development in general, is under genetic control. For instance, dairy farming is an instance of human cultural niche construction that is mediated by cultural processes. There are no genes for dairy farming (i.e., no genes selected specifically for that function). Yet in spite of the fact that dairy farming is not caused by genes and is not a biological adaptation, it has clearly had evolutionary consequences.

Second, even if it were the case that the niche-constructing activities of organisms were under genetic control, it would still not follow that niche construction was a mere effect of the prior selection of these controlling genes. For in many cases, these "controlling genes" have themselves been selected as a result of prior niche-constructed changes in selective environments. For example, it is tempting to assume that the "ultimate" explanation for why earthworms modify soils is that prior selection has furnished them with genes for burrowing, tunneling, exuding mucus, and so forth. However, it is no more than a convention within evolutionary biology that natural selection should be regarded as the ultimate cause of such phenotypic characters. One might equally argue that the cause of the selection pressures that favored earthworm soil-processing adaptations is the prior niche-constructing activities of ancestral worms, without which there would be no soil environment to act as a source of natural selection. The niche construction perspective rejects Mayr's proximate–ultimate causation reasoning, replacing it with the notion of "*reciprocal* causation."

Misleading Metaphors

Lewontin argues that some unfortunate legacies of Darwin and Mendel are misleading metaphors. Mendel's (1983) view of organisms as the manifestation of autonomous internal "factors" with their own laws germinated into a post-Synthesis metaphor in which ontogeny "is seen as an *unfolding* of a form, already latent in the genes" (p. 276). Darwin's view of organisms as passive objects molded by the external force of natural selection encouraged a conception of evolution in which "the environment poses the problem"; the organisms posit "solutions," of which the best is finally "chosen" (p. 276). The metaphor of selection, inspired by the efficacy of artificial selection, continues to encourage a view of organisms as passive objects on which external forces act.

The Niche Construction Revision

The niche construction perspective differs from the conventional perspective in recognizing *two* major adaptive processes in evolution, natural selection *and* niche construction, and two general forms of inheritance in evolution, genetic *and* ecological inheritance (the latter including cultural processes). There are two legacies that organisms inherit from their ancestors, genes and modified environments, incorporating modified selection pressures. Ecological inheritance is not a high-fidelity template-copying system such as genetic inheritance. Instead, organisms transmit to their offspring, and subsequent descendents, physically (and in the case of humans, culturally) altered selective environments, both through actions on their biological and nonbiological environments and by their habitat choices.

The solution to the reference device problem was to change the explanatory reference device. Instead of describing the evolution of organisms relative to natural selection pressures in independent external environments, F. J. Odling-Smee and colleagues (2003) describe evolution relative to the "niches," or organism–environment *interactive* relationships. Because niches are defined by *two-way* interactions between organisms and their environments, this step allowed an "interactionist" theory of evolution to be substituted for the standard externalist account. The niche is a neutral explanatory reference device for evolutionary theory that can capture reciprocal causation without imposing any bias in favor of natural selection and against niche construction, or vice versa.

On the basis of this revised explanatory reference device, all developmental processes that modify the organism–environment relationship are potentially evolutionarily causal. NCT replaces proximate and ultimate causation with "reciprocal causation" and regards the characteristics of organisms as caused by interacting cascades of selection and construction, described elsewhere as "cycles of contingency." From the beginning of life, all organisms have always, in part, modified their selective environments by niche construction, and their ability to do so has always, in part, been a consequence of natural selection.

This recognition of reciprocal causation in evolution also goes a long way toward addressing Lewontin's concerns regarding the misleading metaphors of unfolding developmental programs and the externalism of natural selection. The organism is viewed as both a cause of its own development and a cause of its own selective environment. The key task for any developing organism is the active *regulation* of its inherited "niche," both by responding to its environment and by altering its environment, in ways that keep its personal organism–environment relationship continuously adaptive, for the rest of its life.

Human Cultural Niche Construction and Sociocultural Change

Humans are, of course, one of the species that actively modify their environment; in fact, human beings are the most energetic and wide-ranging species in this respect, continuously affecting their environment through cultural changes that, in turn, modify the selection pressures on their genes. Cultural niche construction refers to evolving sociocultural traits (e.g., dairy farming, cooking, weapons, property rights, discriminate sociality, or educational patterns linked to preferences in family sizes) shaping specific cultural niches that impinge upon the evolution of genetic as well as other cultural traits.

NCT has recently moved into human niche construction that results from cultural (social, economical, etc.) practices that are shown to precipitate the creation of niches. In doing so, NCT has tried to show, both by empirical research and by mathematical modeling, that its theoretical insights and methodological precepts can have an impact on the human and social sciences, from archaeology to sociology. Conversely, support for NCT has very recently come from other quarters, namely the

extended-mind hypothesis and distributed-cognition thesis in the philosophy of mind.

By engaging with the biological role of cultural practices—not themselves biological adaptations—NCT has also corrected and developed further the thesis of gene–culture coevolution. By applying itself to humans, NCT has been able, unlike alternative paradigms like sociobiology, to give to culture (or economic, social, symbolic, etc., practices) its accurate causal role in evolution, explaining correctly how acquired traits can be transmitted. NCT provides a conceptual framework appropriate to the social sciences that is both biologically and culturally based, recognizing the role of human agency whereby humans are, in addition to natural selection, themselves part causes of their own evolution and history. In fact, it has been argued that niche construction as a result of human cultural practices can sometimes be even more effective than gene-based niche construction in directing evolutionary processes and affecting human genes (the human mind, in particular). It is also more rapid. Like biological evolution, cultural processes can be seen as functional solutions to problems posed by the environment, but unlike the former, the latter can be seen to be shortcuts to acquiring adaptive information: This is because human individuals can quickly copy (learn rapidly from) more knowledgeable or experienced others (e.g., when told what to eat or how to avoid danger), "allowing naive individuals to shortcut the many iterations of ontogenetic selection necessary to learn for themselves behavioral patterns appropriate to their environment and thus leapfrog to the functional and already-tested solutions established by others" (Laland, 2011, pp. 196–197). In other words, rapid cultural responses to a culturally modified niche can render genetic reactions redundant.

There are two interesting examples of how NCT may be significant to social-scientific inquiry. One study (Gintis) shows the impact of gene–culture coevolution on the internalization of norms and character virtues and argues that cultural features of the social environment directed the genetic evolution of predispositions, for example, moral cognition. Another study (Shennan) reverses the order of explanation of an important economic phenomenon, wealth inequality, challenging standard social-scientific explanation as well as orthodox evolutionary models.

Finally, one important implication of NCT epistemological proposals is interdisciplinarity: More than one discipline seems to be required since no discipline on its own is adequate to establish cause and effect in cases of gene–culture interaction.

Kevin N. Laland

See also Biology and the Social Sciences; Cultural Evolution; Evolutionary Psychology; Primatology and Social Science Applications; Sociobiology

Further Readings

Gintis, H. (2011). Gene–culture coevolution and the nature of human sociality. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366,* 878–888.

Godfrey-Smith, P. (1996). *Complexity and the function of mind in nature.* Cambridge, England: Cambridge University Press.

Kendal, J. R. (2011). *Cultural niche construction and human learning environments: Investigating sociocultural perspectives* [Special issue]. *Biological Theory, 6*(3), 241–250.

Kendal, J. R., Tehrani, J. J., & Odling-Smee, J. (Eds.). (2011). Theme issue "Human niche construction." *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1566), 785–792. Retrieved from http://rstb.royalsocietypublishing.org/content/366/1566.toc

Laland, K. N., & O'Brien, M. J. (2010). Niche construction theory and archaeology. *Journal of Archaeological Method and Theory, 17,* 303–322.

Laland, K. N., & O'Brien, M. J. (2011). Cultural niche construction: An introduction. *Biological Theory, 6*(3), 191–202.

Laland, K. N., Odling-Smee, J., & Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. *Behavioral and Brain Sciences, 23,* 131–175.

Laland, K. N., Odling-Smee, J., & Feldman, M. W. (2001). Cultural niche construction and human evolution. *Journal of Evolutionary Biology, 14,* 22–33.

Laland, K. N., Odling-Smee, J., & Myles, S. (2010). How culture shaped the human genome: Bringing genetics and the human sciences together. *Nature Reviews, 11,* 137–148.

Laland, K. N., & Sterelny, K. (2006). Seven reasons (not) to neglect niche construction. *Evolution, 60,* 1751–1762.

Lewontin, R. C. (1983). Gene, organism, and environment. In Bendall, D. S. (Ed.), *Evolution from molecules to men* (pp. 273–285). Cambridge, England: Cambridge University Press.

Odling-Smee, F. J. (1988). Niche constructing phenotypes. In H. C. Plotkin (Ed.), *The role of behavior in evolution* (pp. 73–132). Cambridge: MIT Press.

Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). *Niche construction: The neglected process in evolution* (Monographs in Population Biology 37). Princeton, NJ: Princeton University Press.

Shennan, S. (2011). Property and wealth inequality as cultural niche construction. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366,* 918–926.

Further information can be found on a niche construction website: http://lalandlab.st-andrews.ac.uk/niche/index.html

Human Geography, Social Science of

When first confronted with the literature on how geographers construct their world philosophically, the reader is faced with a bewildering set of ostensible alternatives. As a named discipline, geography is an ancient form of intellectual inquiry, predating Greek classicism and its notions of rational thinking. While its Greek meaning—*earth writing*—hints at the importance of geography as a disciplinary practice and its central concern for how humans and earth interact, what is thought of as philosophy in geography has changed radically since the early 20th century, with the last half-century, in particular, resulting in an increasingly conflicted and contradictory set of arguments for how a geographic way of knowing is (and, indeed, should be) constituted and practiced.

Philosophy and Geography

For the most part, philosophy in human geography (or how human geography is thought about) is approached with a keen eye on how people, places, and spaces intersect, interact, and interrelate. Because it is important to capture contemporary geography as an intellectual as well as practiced discipline that is internally differentiated and contested, while at the same time we should understand that knowledge is always partial and practice is often infused with passion, this entry does not attempt to elaborate the entire corpus of knowledge that constitutes contemporary human geographic thought. Rather, it points to what is central from the standpoint of the philosophy of social sciences: the contested and hotly debated nature of diverse ways of knowing within the discipline.

One way of approaching philosophy in human geography is to provide a linear and historical appraisal of how knowledge is built and transformed. There is a commonly accepted narrative of what might be thought of as a patterned sequence to how geographers have come to know the world. In this formulation, the discipline's so-called paradigms—or "isms"—stretch back over time and help define what follows. This way of structuring knowledge is essentially about lumping philosophies into categories that may begin, for example, with environmental determinism in the early 20th century and then flow through possibilism, regionalism, the quantitative revolution, structuralism, realism, humanism, Marxism, feminism, queer geographies, postmodernism, and postcolonialism, to end perhaps with poststructuralist and relational perspectives today.

Earlier Issues and Tendencies

The notion in the early 20th century that the *environment* determines human behavior, resulting in the suggestion that certain peoples and cultures at lower latitudes are enervated by torrid conditions while those in temperate climates rise to become great civilizations, is quickly criticized. This way of thinking was succeeded by the idea that environments are limiting at worst, while many things are possible (i.e., possibilism). In the 1920s, geographers in France and California focused on the importance of *Le Pays*—or the region—as a unifying motif for how we practice living, but this idiographic approach with its focus on the uniqueness of place was challenged by a push for explanation, generalizability, and scientific objectivity as positivism swept through the social sciences from the 1950s onward.

Very quickly, the notion of scientific objectivity was challenged by humanism and feminism, joining a strand of anarchism and Marxism that was present in human geography since the mid 19th century with the work of Peter Kropotkin. By the 1970s, postcolonialism refocused intellectual narratives from the center of colonial power to the so-called geographic margins (former colonial spaces). Humanism, feminism, queer studies, and postcolonialism also contributed to a growing discussion of political identities that changed the focus of the geographical imaginary toward subaltern, psycho-, and corporeal geographies. Poststructuralism in geography pushes the study of the politics of identity and subjectivity

to include the study of representations and representational and nonrepresentational politics.

The "isms" suggest abstract knowledge that is extracted and simplified from a very complex set of interactions between people, places, and intellectual movements. It is common with this kind of delineation for commentators to smooth out and generalize the connections between different philosophical approaches or sensibilities. Too often, the contested nature of the world and our knowledge of it are neglected by supplying a relatively linear set of approaches melding into each other and ending with a particular writer's preferred way of knowing. Readers often fail to grasp the contested nature of the discipline and regard the approaches as "pick-and-mix" alternatives rather than recognizing the tensions between those who adopt different philosophical positions. Tensions often revolve around what a set of philosophies and theories proposes as a basis for geographic knowledge and how practical those philosophies and theories are in delivering that knowledge. Indeed, the debates between the particular (idiographic) and the general (nomothetic) that populated the pages of academic geography in the 1950s return in different forms throughout the last half-century with critiques of metanarratives, arguments between materialist and idealist approaches, local and global perspectives, discussions about the merits of humanistic, poststructural, and relativistic approaches, and so on. Because the context of discussions changes at different times and in different places, the point is not just about what is contested and resolved but that contestation is creatively adopted and used to philosophically propel geographical ways of knowing and understanding.

Current Approaches

Today, human geographers are taking on very interesting philosophical questions of how contestation does not have to necessarily be understood negatively. Instead, drawing on the "immanent" and "affirmative" philosophies of thinkers such as Baruch Spinoza and, more recently, Gilles Deleuze and Antonio Negri and intersecting them with the insights, concerns, and critiques of more traditional intellectual streams in geography—such as feminism, ecology, and Marxism—geographers have begun challenging how staple geographical unifying concepts, such as scale, environment, landscape, time, and place and space, and wider social concepts such as class, race, gender, and politics are approached, thought about, communicated, and consumed. "Immanence" muddies the connections between things internal, external, outside, part of, and beyond while simultaneously mandating connections with the material and corporeal. "Affirmative" philosophies are about disambiguation and creating contexts for positive and constructive practices for people as well as the nonhuman and the more than human. We are deeply embedded in our very particular geographies, and those geographies are variously connected to larger communities and economies to the extent that a complex web of different relations circumscribes our daily embodiments, practices, and politics. Human geography today grapples with some of the pithy philosophical questions that arise from these relations in fluid and uncompromising ways and pays keen attention to how their material and intellectual intersections and entanglements continually affect and produce different people, places, and spaces.

Stuart C. Aitken and Giorgio Hadi Curti

See also Paradigms of Social Science; Positivism, History of; Postcolonial Studies; Postmodernism

Further Readings

Aitken, S. C., & Valentine, G. (2006). *Approaches to human geography: Philosophies, people, practices.* Thousand Oaks, CA: Sage.

Del Casino, V. (2009). *Critical introductions to geography: Social geography.* Oxford, England: Wiley-Blackwell.

Harvey, D. (2001). *Spaces of capital: Towards a critical geography.* London, England: Routledge.

Henderson, G., & Waterstone, M. (2009). *Geographic thought: A praxis perspective.* New York, NY: Routledge.

Massey, D. (2005). *For space.* Thousand Oaks, CA: Sage.

Human–Machine Interaction

Philosophy has always sought to pose the larger questions in life and provide answers. Two of the most important of these questions are (1) What, if anything, differentiates human beings from all other

forms of life? and (2) What does it mean to be human? The latter question has recently been extended to include concerns regarding the distinction between humans and machines, especially when these two interact. These questions are addressed in this entry. Its primary frame of reference is the consideration of humans and their interaction with machines.

Human Beings as Differentiated From All Other Forms of Life

Many differing characteristics have been championed as key elements that differentiate human beings from all other orders of life. In itself, the fact that this question is posed so often indicates how desperately humans have sought to distinguish and divorce themselves from the rest of the animal world. Among the candidates offered are the utilization of language, the capabilities rendered by the opposable thumb, the capacity for an upright gait, and even the form of face-to-face sexual intercourse, which is very rarely practiced outside the human species. Such characteristics are certainly attractive as potential causal explanations, but whether they are, individually or collectively, open to empirical resolution is itself an interesting conundrum. What is offered in this entry is an account founded on the use of advanced forms of tools, specifically machines. As will become evident, we seek also to distinguish "mere" tools from externally powered machine systems.

Evolution is typically associated with Charles Darwin's theory, in which each individual organism adapts to its ambient environment and certain associated and subsequently transmittable advantages are "selected" as a function of experienced, contextually contingent pressures. But what of the process of selection in our contemporary human species? What are the pressures exerted by our environment that currently surrounds us? Most likely, you are reading this tract in book form or via some technological appendage in a designed and "artificial" environment that has been created by human beings who have preceded you. What are the "natural" constraints you are now experiencing? Our answer is that such constraints are now predominantly self-determined. That is, humans have conceived, created, and constructed the main fabric of the modern world in which we live. Thus, although we acknowledge and recognize that there is quite a spectrum of tool use in the animal kingdom and indeed that most living systems look to optimize their own living conditions as far as it is within their control, it is only we humans who occupy a dominantly self-manufactured environment. More formally, specific, species-produced *orthotics,* which then themselves co-adapt, are confined solely to human beings. That co-adaptation occurs at a frequency that is derived from the integration of the respective timescales of change as represented by variation rate in the organism (human) and the orthotic (machine) themselves. Peter A. Hancock and Gabriella Hancock have termed this characteristic *the self-symbiotic species*. This is perhaps *the* dimension that makes human beings unique.

Now one of the great conflicts of the late 19th century can be reconciled. Humans are indeed animals in that they possess the fundamental structure and functions common to all animal life. Thus, Darwin was correct. Yet we are not *only* animals. We have been involved, sui generis, in the creation of a new form of hybrid species in which we are ever more progressively conjoining with the product of our own minds. We are certainly not gods, but we have used what the English mathematician and occultist John Dee (1581) called *thaumaturgike* (he referred to it as a low form of magic) to elevate ourselves beyond any other living system (at least any that is currently known to us).

What It Is and What It Will Be to Be Human

Although we must be careful to distinguish between tools and machines, we can, as a general statement, propose that tools created humans as much as humans created tools. (As Peter Hancock explains, our primary differentiation here is that a machine derives its motive power from a source beyond its immediate user, as compared with a tool, which is directly powered by the individual who wields it.) Furthermore, from a topical examination of brain architecture, we can sequentially conclude that a tool (and its evolutionary offspring, the machine) can be thought of as both a cause and an effect of the imagination. We are *self-symbiotic,* first with tools but now with the machines that we create. In case one doubts this assertion, it is quite possible to induce significant distress in any of our modern

generation simply by parting them from their handheld computational and communication devices for even just one day! But this process of coevolution is far from finished, and indeed one can reasonably argue that it has barely begun. In light of this development, we now have to ask rather difficult philosophical questions such as "What are the boundaries of the human condition?" It is very evident that we cannot abandon our technology and remain the current incarnation of the present species. Yet technological evolution progresses at a dissociatingly fast rate of development compared with intrinsic human change. Thus, our coevolutionary path seems disproportionately driven by profit-driven, technical advances.

Our linkage with our machines is also becoming physically as well as cognitively more intimate. Many people now have in-dwelling medical diagnostic and therapeutic devices. It will not be long before such implants are primarily elective rather than medically obligatory. What will it mean to be human when we are progressively more machine in composition? At this juncture, the questions of science fiction and philosophical contemplation become intimately and indeed alarmingly related. The issues of ownership, privacy, responsibility, and legal culpability are all immediately brought into play when hardware and software physically enter humanware. And with our viral capitalistic structure, the violent shadow of profit is also sure to enter the equation associated with such deliberations. However, we have been instructed that the purpose of philosophy is not simply to study the world but to change it. Thus, we have to conclude here by asking questions not of what is and what may well be but rather what *should* be. Thus, our conclusion is certainly a value judgment, but in such times, statements of value are mandated.

We are in particular danger of associating, attaching, embedding, and enabling an insufficiently capable biological system (the human being) to an ever more powerful and evidently destructive capacity (the ascendingly complex and interrelated systems of global technology). Our media are decorated with the failures that accrue when the fallible human is overwhelmed by the demands of the voracious machine. If human error is an expression of passive malevolence, we also live with the specter of actively malevolent humans let loose with portable and awesomely destructive weapons. How then can we regulate this emerging symbiosis? Can we inculcate morality (e.g., safety) into technology by steps of pure design? It seems evident that we are in a race to establish the next state of punctuated equilibrium in this symbiotic evolution before we destroy the very fabric of the environment that sustains us. Some observers find reason to believe that we shall fail in this endeavor.

Human–Machine Interaction: Approaches

Approaches for research (and the theories that drive such science) should therefore understand this unique relationship between humans and the technology they use, so that future design efforts and training methods foster more effective human–machine interaction (HMI) rather than promote discord. To this end, research in this domain has been often multidisciplinary in nature; most commonly, a two-pronged approach combining both behavioral and computer science. Design and industrial engineering provide insight into the mechanical perspective of the human–machine dyad, while cognitive psychology and ergonomics attempt to explain and predict the physical and psychological reactions and performance of the human operator. Human factors psychologists (i.e., those working in the multidisciplinary Human Factors Science or ergonomics), however, merge each of these respective disciplines to work on the HMI as its own complete system, investigating its effectiveness, usability, interface structure, and the like.

Current Trends

Some of the major areas of interest in HMI today are augmented reality, individuation/customization, and embodied cognitive agents. Augmented reality research looks to create an interface whereby the human is able to perceive an environment with overlaid information that would normally be available to the senses. For example, in the Google Glasses project a person wears a set of transparent glasses to successfully interact with the world in front of her; projected onto the glasses themselves, however, is information about the person's surroundings (i.e., a GPS screen provides her with directions on where to walk, pop-ups alert her that there is a coffee shop near her current location, etc.). But, to come

to the second major area mentioned above, all such technologies are rapidly becoming more attuned to the individual and less designed for inflexible mass consumption. In particular, with the overwhelming number of personal devices now available, users are able to choose a number of personalized settings: backgrounds, schemes, ringtones, covers, and so on. Moreover, some devices have been programmed to respond only to the sound of an authorized user's voice. We anticipate that such individual customization or individuation will burgeon greatly in the near future.

Other contemporary thrusts have featured embodied cognitive agents through which some machines are now able to reciprocate individuated types of communication. An embodied cognitive agent is a machine or program displaying a limited amount of artificial intelligence (AI), which is anthropomorphized to an extent so as to give the human user a sense of social interaction with the technology. For example, Siri is an application for the iPhone that works as an intelligent personal assistant; "she" is able to vocalize and talk with the user, ensuring that "her" findings are indeed what the user wanted.

Major Players: Past and Present

Throughout the 20th century, scientists from a number of disciplines contributed significantly to the theories and research underlying HMI. For example, Alphonse Chapanis of Johns Hopkins University is generally considered the father of this area, in particular ergonomics, at least in the United States. Another early luminary was Paul M. Fitts Jr., whose most preeminent contributions concern his law of motor performance and his study of piloting error. As one of the founding fathers of aviation psychology, he was also interested in maximizing the efficiency of human movement necessary to interface with machine (airplane) controls. In respect to decision making as related to HMI, Herbert Simon, educated at the University of Chicago and a Nobel Prize winner in economics, made a significant contribution to HMI research with his work on AI. His work with Allen Newell on the Logic Theory Machine and the General Problem Solver helped further the capabilities of machines to perform more complex, decision-making tasks. Donald Norman, a graduate of MIT and the University of Pennsylvania, has advanced HMI research with his concept of "user-centered design," which dictates that engineers and designers primarily focus on designing a machine based on the human user's needs rather than on convenient engineering or aesthetics. Last, Ivan Sutherland (who was educated at CalTech and MIT) was one of the pioneering scientists at the forefront of the development of the graphical user interface. His invention, Sketchpad, was the prototype for software on which most modern-day personal computer systems are now based. Sutherland was also one of the computer scientists who contributed to the development of the Internet, arguably the invention that has most radically altered HMI since the advent of technology.

P. A. Hancock and G. M. Hancock

See also Agent-Based Modeling and Simulation in the Social Sciences; Artificial Intelligence; Evolutionary Psychology; Machine Consciousness and Autonomous Agents; Social Studies of Science and Technology; Technological Convergence; Technoscience and Society; Transhumanism and Human Enhancement

Further Readings

Boy, G. A. (Ed.). (2011). *The handbook of human-machine interaction: A human-centered design approach*. Farnham, England: Ashgate.

Hancock, P. A. (2009). *Mind, machine and morality.* Chichester, England: Ashgate.

Hancock, P. A., & Drury, C. G. (2011). Does human factors/ergonomics contribute to the quality of life? *Theoretical Issues in Ergonomic Science, 12,* 416–426.

Hoc, J.-M. (2000). From human-machine interaction to human-machine cooperation. *Ergonomics, 43,* 833–843.

Mazlich, B. (1993). *The fourth discontinuity: The co-evolution of humans and machines* (10th ed.). New Haven, CT: Yale University Press.

Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics, 38*(8), 114–117.

Moravec, H. (1988). *Mind children: The future of robot and human intelligence.* Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

Norman, D., & Draper, S. (Eds.). (1986). *User centered system design: New perspectives on human-computer interaction.* Lawrence Erlbaum, NJ: CRC Press.

Zetterberg, J. P. (1980). The mistaking of "the mathematicks" for magic in Tudor and Stuart England. *Sixteenth Century Journal, 11*(1), 83–97.

Hypothetico-Deductivism

Hypothetico-deductivism (H-D) is an account of the confirmation or testing of (scientific) theories, typically by empirical (observational) evidence. Some philosophers, most notably Karl Popper, have offered H-D as an alternative to inductive accounts of confirmation. H-D's core claim is that a theory or a hypothesis is confirmed or tested by testing its deductive observational consequences. Arguably, it is the version of confirmation that is still most prevalent among working scientists. However, attempts to provide a rigorous formulation of H-D have led to a series of technical problems that have led most leading philosophers of science to reject H-D. Recently, it has been argued that these formal problems are not fatal but that H-D suffers serious philosophical shortcomings, namely, that it does not sufficiently account for inductive support and does not provide an account of the confirmation of statistical theories. This suggests that H-D can at best give a partial account of confirmation.

The Formal Accounts of H-D

Although H-D has been attributed to the 19th-century scientist and philosopher of science William Whewell, this attribution, and especially the related attribution to Whewell of a wholesale rejection of induction, is questionable. Rather, it is with 20th-century writers such as Ernest Nagel, Karl Popper, and Carl Hempel, coming after the development of modern logic, that the distinction between H-D and induction and the attempt to formulate formal accounts of H-D come into sharp focus.

At its simplest, H-D is the claim that (observational) evidence E confirms hypothesis H if and only if H logically entails E. A key objection to this simple version is that many hypotheses deal with unobservable entities and hence have no direct bearing on observation. Consider, for instance, the claim that all electrons are negatively charged. Rather, such hypotheses when supplemented with various background auxiliary assumptions and statements of initial conditions entail observational consequences. This leads to more complicated versions of H-D: For instance, evidence E confirms hypothesis H relative to background information B if the conjunction of H and B entails E, but B alone does not entail E. This needed complication brings additional worries of its own. It may be argued that knowing that some H is confirmed relative to some B by observational evidence E does not tell us if H has itself been confirmed. Some critics want an account of how we get from relative confirmation to simple confirmation, their worry presumably being that for any hypotheses H there will always be some observational evidence E that will confirm H relative to some totally crazy and/or ad hoc background assumption B. Others, pressing what has come to be known as holistic accounts of confirmation, have questioned thus: If it is B and H that together entail E, why should we take the confirmation to accrue to H rather than to B? Alternatively, why not simply say that E confirms the combination of H and B? Analogously, some have raised similar questions about disconfirmation: Where H and B together entail E, but we find that E is false, why should we say that it is H rather than B—or, alternatively, the combination of H and B—that has been disconfirmed?

For the time being, we can put aside these questions and concentrate on the simplest version of H-D so as to appreciate the formal problems that have led philosophers of science to reject H-D.

Two Formal Problems

Rather than survey all the formal problems, we shall briefly consider the two main ones.

Tacking by Conjunction

If H entails E, then for arbitrary H′ the conjunction of H′ and H entails E. So if deductive entailment is sufficient for confirmation, then where E confirms H it also confirms the conjunction of H and arbitrary H′. So finding that the first planet of the solar system travels in an elliptical orbit around the sun confirms not simply that all the planets of the solar system travel in elliptical orbits around the sun but also the conjunctive hypothesis that all the planets of the solar system travel in elliptical orbits around the sun and all the planets of the solar system are made of rubber.

Tacking by Disjunction

If hypothesis H entails E and E′ is some arbitrary true observational claim that has nothing to do with H, then H entails the true observational claim E or E′.

So for any hypothesis H there will always be some true observational claim that confirms H. Thus, "All the planets of the solar system are made of rubber" is confirmed by the true observation claim "The first planet of the solar system is made of rubber, or the first planet of the solar system travels in an elliptical orbit around the sun."

Recently, it has been argued that these and other formal problems may be avoided by recasting H-D in terms of natural axiomatizations of theories. The idea is that a theory has certain canonical (natural) representations, and only those axioms of a canonical representation that are needed in the derivation of E are confirmed by E. So where H entails E, it is true that the conjunction of H and H′ also entails E, but only H is needed in the derivation of E, so only H is confirmed by E. That disposes of the tacking-by-conjunction objection.

To get rid of the tacking-by-disjunction objection, we can demand that for hypothetico-deductive confirmation the relevant E should not simply be a logical consequence of the relevant H but should be part of its content. The idea here is to say that while "All the planets of the solar system are made of rubber" logically entails both "The first planet of the solar system is made of rubber" and "The first planet of the solar system is made of rubber, or the first planet of the solar system travels in an elliptical orbit around the sun," only the former counts as part of the content of "All the planets of the solar system are made of rubber." Thus, we might claim that E only hypothetico-deductively confirms axiom H of a theory if E is part of the content of H.

Philosophical Problems

Leaving aside these formal problems, there are a number of deep philosophical problems that suggest that even if there is something to H-D it should not be presented as providing a necessary condition for confirmation.

The first philosophical problem is that canonical formulations of H-D do not allow for inductive support. Consider that of the two claims "All the planets of the solar system travel in elliptical orbits around the sun" and "The seventh planet of the solar system travels in elliptical orbits around the sun," only the former entails "The first six planets of the solar system travel in elliptical orbits around the sun." So where H-D provides *necessary* and sufficient conditions for confirmation, it has the consequence that the true claim "The first six planets of the solar system travel in elliptical orbits around the sun" confirms "All the planets of the solar system travel in elliptical orbits around the sun" but does not confirm "The seventh planet of the solar system travels in elliptical orbits around the sun." In other words, H-D taken as providing necessary and sufficient conditions for confirmation actually precludes inductive support. While Popper may have taken this as a virtue of H-D, most would take the view that if confirmation of a theory gives no reason to think that the theory will work in the next case, then we have little reason for caring whether or not our theories are confirmed.

The second major philosophical problem is that a statistical hypothesis, for instance, that the probability of a birth in England being of a male child is 50% or that the half-life of radon isotope 222 is 3.8 days, has no *deductive* observational consequences, and hence, if H-D is correct, there is no observational evidence that can confirm the hypothesis.

In light of these objections, H-D should be recast as giving only sufficient and not necessary conditions for confirmation. This would allow that H-D could be combined with a Hempelian transmission condition, such as "If E confirms H and H′ is part of the content of H, then E confirms H′." This would allow for inductive confirmation.

Bayesian Confirmation

To allow for confirmation of probabilistic hypotheses, one might relax the required relationship between E and H. Thus, instead of demanding that H entail E, we might simply demand that H make E more likely. So while "The half-life of radon isotope 222 is 3.8 days" does not logically entail that "Any large ensemble of radon isotope 222 will show radioactive decay when measured with a Geiger counter," it does make that claim highly likely. Put in terms of probability theory, the condition that H makes E more likely amounts to the requirement that $P(E/H) > (E)$; in other words, the probability of E given H is higher than the probability of E by itself. Bayesians, following the lead of Carnap, have long claimed that E confirms H if and only if $P(H/E) > P(H)$. What has often gone unrecognized is that that claim is equivalent to the claim that E confirms H if and only if $P(E/H) > (E)$, and hence, Bayesian

confirmation is simply a weakening of the core demand for confirmation laid done by H-D.

It is tempting, and now very common among philosophers of science, to conclude that H-D should be simply abandoned in favor of Bayesian accounts of confirmation. This suggestion itself meets with both formal and philosophical objections. One formal objection is that Bayesianism in its current forms is actually subject to both the tacking objections mentioned above that have been promoted against H-D. The main philosophical objection is that Bayesian accounts rely on the arguably suspect notion of prior probabilities. Given a coin, before any observations, I may have a prior probability of its coming up heads on a given throw of .5. But what is my prior probability, separate from any observation or, for that matter, separate from any theoretical knowledge, for the claim that the first planet of the solar system has an elliptical orbit?

Conclusion

Because of its failure to account for the confirmation of probabilistic hypotheses, H-D, even when supplemented with a Hempelian transmission condition allowing for inductive support, cannot be taken as the whole account of confirmation. However, it should be recognized as giving a sufficient account of confirmation, one not encumbered by fanciful assumptions about wholesale distributions of prior probabilities. Moreover, it should be recognized that it is in practice the way most scientists actually seek to confirm their theories. That is to say, scientists will recognize as good practice the process of (a) giving clear, nonarbitrary formulations of their theories; (b) teasing out observational consequences from those theory formulations; (c) checking those consequences for truth; and (d) raising their confidence in the truth of the parts of the theory involved in teasing out those observational consequences if those observational consequences are in fact borne out.

Ken Gemes

See also Bayesianism, Recent Uses of; Covering-Law Model; Induction and Confirmation; Pessimistic Induction; Popper's Philosophy of Science; Probability; Scientific Method

Further Readings

Gemes, K. (1998). Hypothetico-deductivism: The current state of play; the criterion of empirical significance endgame. *Erkenntnis, 49,* 1–20.

Glymour, C. (1980). Discussion: Hypothetico-deductivism is hopeless. *Philosophy of Science, 47,* 322–325.

Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science.* New York, NY: Free Press.

Nagel, E. (1979). *The structure of science: Problems in the logic of scientific explanation.* Indianapolis, IN: Hackett.

IDEALISM

The concept of *idealism* has a long and complex history. The first half of this entry traces the outline of this history. The second half of the entry focuses on idealism in the social sciences.

Definition

Idealism is a family of philosophical theories according to which ultimate reality is ideal and transcendent or a function of consciousness or reason. Idealism is also the view that human beings should strive to live under the influence of self-chosen ideals. Finally, idealism can be a view of historical change as being the result of mainly ideas, plans, visions, or interpretations. In general, idealists tend to emphasize activity and freedom while being opposed to materialism and determinism. Many idealists are therefore interested in human subjectivity and its status. In discussions of art, idealists valorize creativity and imagination over faithful copying of nature (or realism).

History of Idealism

Ontological Idealism

The word *idealism* is derived from the Greek words *idea* (ἰδέα) and *eidos* (εἶδος), which appear as early as in the works of Homer. The term *eidos/idea* (a derivative of the verb *to see* or *to look*) can mean "image," "shape," "look," "kind," or "species." In Plato, who is the first to employ this term for philosophical purposes, "idea" designates the unchanging, transcendent, and supersensible forms of things—a world of pure essences in relation to which the sensible world is ontologically inferior and no more than a reflection or imitation.

In dialogues such as *Phaedo, Phaedrus,* and *Politeia*, Plato arrives at this view, which by focusing on what constitutes reality as such can be called ontological idealism, from many different directions. In order to tell that a thing is of a certain nature, it is necessary to know something that cannot be derived from contemplating the particular thing itself. In order to know that *this* is a horse, one must have knowledge of what being a horse is, for otherwise one would not be able to identify *this* as a horse. One must know something about what all horses in the world have in common—the conditions under which something may be said to be a horse. Once one arrives at the thought that these conditions cannot be arbitrary and changing and destructible (since, if all horses disappeared, what it is to be a horse must not be possible to touch), then the conclusion at hand is that every way a thing might be must be determined by essences that exist beyond the empirical world, in a world of perfection that will never change.

In Plato and in the Platonic tradition, idealism is the view that reality is supersensible, composed of essences or immaterial (and hence indestructible) forms that human beings can only cognize through thinking, in particular the kind of reflective thinking that Plato calls dialectics. As Platonism started to influence Christianity via Plotinus, Augustine,

Origen, and others, idealism became a crucial component in the attempt to distinguish between good and evil, which was often associated with the distinction between the transcendent and the immanent or between intelligible form and sensible matter. In many Neo-Platonist schools (and already in Plato), idealism also involved distinguishing sharply between the immaterial soul and the material body and thinking that the world is essentially One, a world soul, chained to matter but capable of liberating itself to its original source in pure light, *nous,* or the Good. In Plotinus, in particular, idealism formed the basis of a type of mysticism that called for the fusion of the human soul with the One or the absolute. Thus, idealism was sharply opposed to all views, Epicureanism in particular, that emphasized the material dimension of existence and argued for its ontological priority.

Epistemological Idealism

In the 17th and 18th centuries, with the rise of empiricism and a renewed interest in questions of epistemology, the term *idealism* attained new meanings. Idealism could now mean, as in Bishop Berkeley, that only mental content is real, or it could, as in David Hume, take a more skeptical form and imply that, whereas one must conjecture that a mind-independent reality exists, humans can only be aware of their (internal) impressions and ideas. With Berkeley's *esse est percipi,* the world becomes identical with the thought or representation we have of it. With Hume's skepticism, on the other hand, we can never know with certainty that our representations match any mind-independent reality. The best we can do is to continue acting as though they do. Idealism, if taken to its extreme in Berkeley, may offer a complete ontology and even a new, anti-materialistic worldview. In Hume, however, it threatens to enclose the perceiving and thinking subject inside the circle of its own representations. Idealism, on these views, is contrasted with realism, the belief (ontologically) that mind-independent entities exist and (epistemologically) that we can know them.

Later, Immanuel Kant's transcendental idealism aims to overcome Hume's skepticism while at the same time avoiding Berkeley's equation of mental content with reality. According to Kant, we need to distinguish between the order of things as they exist independently of any possible human cognition and the order of things as they exist for human beings endowed with certain objectivating capacities. We can know whichever entities can be presented to us in accordance with the necessary and universal conditions for representing them. According to Kant, such conditions are space and time as well as certain functions of unity in judgment called categories. The fact that we cannot know things as they are in themselves (qua "*noumena*") but only as they appear to us (qua "*phenomena*") via our schemata is not tantamount to a disastrous epistemological defeat. On the contrary, since it would make no sense to claim to know anything outside of the conditions whereby it can be known, we have reason to be perfectly satisfied with knowledge of appearances. On Kant's view, objective knowledge arises as the result of the correct interplay between conceptual structuring capacities and the constraining impact the world exerts on our senses.

Absolute Idealism

In subsequent phases of German Idealism and, in particular, G. W. F. Hegel's absolute idealism, the Kantian distinction between the order of appearances and the order of things in themselves is vigorously attacked. According to Hegel, it makes no sense to postulate the existence of anything outside the conceptual understanding we already possess. Rather, such understanding is an element of spirit (*Geist*), which for Hegel is a totality without any outside. In Hegel's system, spirit's conceptual self-understanding, including the self-alienation and self-appropriation of spirit, forms the basis for conceiving of human life and history as partaking in a rational structure that develops teleologically. Thus, history itself becomes the unfolding of spirit—the long and arduous process whereby spirit comes to know itself in full freedom.

On a traditional interpretation of Hegel, spirit is God, and everything belongs to and becomes intelligible in light of God's own presence and development in the world. On more recent interpretations, however, spirit is essentially communal, a collective form of self-reflection expressed in art, religion, and philosophy and embodied in the institutions of the modern state. While different in most respects, both interpretations take Hegel's absolute idealism to entail that human history develops along rational lines and that the categories determining the logic of

this development as well as the shape of the modern state can be reconstructed in a speculative philosophy of right. Thus, as Hegel puts it in the preface to the *Elements of the Philosophy of Right,* "What is rational is actual; and what is actual is rational." Idealism in Hegel's account affirms the status quo. It demonstrates that whatever exists is rational, thereby making it possible to reconcile oneself to reality.

In the first half of the 19th century, idealism takes many forms, ranging from Hegel's pan-rationalism to F. W. J. Schelling's system of absolute identity and Arthur Schopenhauer's philosophy of the will. In all of these systems, the material world is ontologically downgraded and seen either as an expression of the absolute (Hegel, Schelling) or as an illusion (Schopenhauer).

In the latter half of the 19th century, the rise to cultural prominence of natural and social science conspired with secularization and modernization to undermine the credibility of absolute idealism. Karl Marx's critique of Hegel is a case in point. According to Marx, Hegel wrongly sought to derive the structure of the modern state from the pure categories of dialectical philosophy. On Marx's view, a materialist account, which refers to the actual social relations between agents in a particular social formation with its specific mode of production, is better able to understand the nature of the modern state than any a priori reconstruction. Rather than Hegel's theoretical approach, which considers social life from the vantage point of the philosophical system, one needs to take into consideration the ways in which agents actively and collectively engage with social reality—what Marx called *praxis*.

British Idealism

Reacting to a long-standing British tradition of empiricism, in the second half of the 19th century and well into the early decades of the 20th, philosophers such as T. H. Green (1836–1882), F. H. Bradley (1846–1924), B. Bosanquet (1848–1923), J. M. E. McTaggart (1866–1925), and H. H. Joachim (1868–1938) developed a species of absolute idealism that has come to be known as British idealism. Thinkers associated with this movement viewed the notion that philosophy should concentrate on the epistemological question of how the mind is able to represent a mind-independent reality as a recipe for skepticism. In the interest of developing a nonskeptical understanding of reality, they rejected the subject–object duality and formulated a *coherence theory of truth* according to which truth arises from a coherent system of beliefs rather than some relation between belief and external reality.

Ultimately, the British idealists believed in the ontological primacy of a self-organizing, self-sufficient sphere of absolute or ideal being. Drawing on Hegel (yet without his peculiar form of dialectical logic), on their account the highest task of reason consisted in conceptually articulating what they thought of as the structure of the absolute.

In contributions to social philosophy and ethics, the British idealists typically emphasized collectivity and social unity over individualism and social atomism. While far from espousing reactionary forms of authoritarianism, they criticized liberalism and tended to lament what they saw as the fragmentation of modern society.

While dominant for some decades and a rare example of the extended Continental influence on British philosophy, mainly at Oxford, British idealism was vehemently attacked and discarded by the subsequent generation of logical positivists and linguistically oriented philosophers. The rise of analytic philosophy dealt a death blow to British idealism (despite some unsuccessful rearguard battles fought by Collingwood and later on by the Hegelian G. R. G. Mure). In the hands of thinkers like Bertrand Russell and G. E. Moore, the British idealists had to be left behind for a new and purportedly more rigorous philosophy to emerge. Since the late 1930s, British idealism has had little impact on contemporary philosophy. Recently, there is some renewed, if somewhat isolated, interest in their thought on metaphysics but quite a burgeoning research activity as far as their political philosophy is concerned.

Idealism in the 20th Century

Idealism continued to play an important role in 20th-century philosophy. In Edmund Husserl's transcendental phenomenology, the immediate givenness of intentional (mental) content is appealed to in order to explore the essential features of reality. Since what we can call real must be possible to present for consciousness, it can be studied from within the immanence of what Husserl calls "pure consciousness"—that is, content as it presents itself

for consciousness without any reference to actual existence.

Phenomenology explores the meaning of the phenomenon as it presents itself to consciousness. In the hands of Martin Heidegger, Jean-Paul Sartre, and many other prominent theorists, it becomes a powerful and nonreductive approach to the study of both subjective and intersubjective phenomena. A crucial part of the attraction of phenomenology is its apparent ability to refute relativism and ground science in an account of rationality.

Much of 20th-century philosophy has rejected the idea that human beings have unmediated access to an objective reality. In phenomenology, and also in pragmatism, poststructuralism, hermeneutics, critical theory, and much of post-Wittgensteinian philosophy of language, the fundamental claim has been that knowledge and understanding are mediated by essentially public structures like language, practices, horizons of pregiven meaning, or ideology. Similarly, in analytic philosophy, so-called anti-realists have held that truth must be a function of warranted assertibility. As opposed to realists, who argue that meaningful propositions have an objective truth-value independently of any possible verification or justification, anti-realists claim that what we can call truth must always be possible to justify. Truth is therefore relative to our actual language-games. Although not everyone has concluded from considerations such as these that knowledge or understanding is therefore restricted, or that it cannot attain objectivity, the emphasis on mediation means that idealism has played a major role in recent philosophical endeavors.

Idealism in Aesthetics

There are several reasons why the term *idealism* has been important in thinking about art. In classicism, and also in rationalism, a beautiful work of art instantiates certain ideal or formal properties that the theoretician seeks to disclose. The successful work of art thus strives to approximate an idea of formal perfection.

Another strain of idealist aesthetics, influenced by Plato, emphasizes transcendence. The successful work of art intimates the eternal ideas existing behind phenomena. The work of art does this either by presenting these ideas directly or by alluding to them. In both cases, aesthetic experience is supposed to lift the beholder out of everyday empirical existence. In the most Platonic accounts, such experiences are supposed to liberate the soul from enslavement in the body.

A third strain of idealist aesthetics, often but not necessarily associated with Romanticism, points to expressive freedom, creativity, originality, and imagination as sources of art. As opposed to realist views of art, which explore the senses in which a work of art can be said to be *about* something, such theories view art as affirming an account of the human subject as free and autonomous. Art becomes a privileged vehicle for manifesting freedom. In Kant's theory of genius, the Romantics found reason to celebrate the artist as an exceptional being, sometimes exempt from bourgeois conventions and everyday morality.

Idealism in the Social Sciences

Idealism in the social sciences is generally the view that ideas and self-interpretations can play an explanatory role when accounting for social phenomena and change. Opposed to this view is materialism, which for explanatory purposes emphasizes the economy, power, material structure, or psychological hardwiring. Max Weber's study of the rise of capitalism is a good example of an idealist approach to social explanation. According to Weber in *The Protestant Ethic and the Spirit of Capitalism,* Calvinism, with its deeply felt need to prove oneself worthy of salvation in the eyes of God, created a mind-set focused on investment and frugality. Although other factors had to be in place, capitalism was largely the result of the implementation of this mind-set. Weber's argument entails that what agents think, feel, and say about themselves influence and often cause their actions. Indeed, radical social change, such as the rise of capitalism, may come about as the direct result of agents having certain beliefs and desires; thus, what people believe and desire can be enormously consequential. Weber bolsters this view by developing a theory of intentional action. Actions, Weber argues, are the result of intentions; their aim is to actualize the specific purposes sought by the agent.

Other idealist accounts of social life include those of Peter Winch and certain versions of discourse analysis. According to Winch, who was deeply influenced by the later Wittgenstein, social behavior is essentially rule governed. To act is to follow a rule that is shared among all members of a particular

community or life form. Discourse analysis, by contrast, is based on the view that people's actions, identities, and self-interpretation are constituted via discourse. Following structuralism, which can also be said to represent an idealist orientation, discourse analysts, like the contemporary social theorist Ernesto Laclau, argue that discourse is structured around an underlying set of oppositions and something is what it is only through its differential relations to something else.

The most obvious objection to idealisms of these sorts is that they underestimate the causal and explanatory power of events and occurrences in the materially existing world and that they are blind to power, class, and other factors external to the self-understanding of agents. As Marx famously summarizes this objection in *The German Ideology,* "It is not consciousness that determines life, but life that determines consciousness."

Idealism is not only juxtaposed to materialism. In debates in political science, one sometimes hears of idealist versus realist approaches to international relations. While realist approaches see power as the underlying factor in understanding how nations come to set and implement national and international priorities and think of international politics as an essentially anarchistic realm of struggle for security and power, idealists see the existence of a self-contained world of laws and rules as essential. The idea behind idealism is that successful relations between nations do not rest exclusively on a balance of power but on the explicit rejection of force grounded in self-enforced rules of behavior. Idealism in this regard is often associated with the political views of Kant, who argued that relations between states should ideally be governed by laws reflecting the general will of all states. Realism, by contrast, tends to be associated with Hobbes, who in the *Leviathan* argued that, rather than legitimacy grounded in law, the fundamental task of all political arrangements is stability, the provision of order and security.

Idealism, Personhood, and the Good

An idealist is sometimes characterized as a person who acts on the basis of ideals—that is, on some conception of the good. Unlike realists, who tend to be pessimistic or stoically detached, idealists in this sense are *optimistic,* believing that their goals will be achieved or at least that it is worth striving to achieve them even when the outcome of doing so is uncertain. The idealist can have this faith because he or she thinks that being committed to certain ideals is good independently of what their actualization may lead to. Idealism in this sense is related to a vision of moral personhood according to which our identity, our sense of selfhood, is a function of the kinds of commitments we make and the clarity and seriousness with which they are articulated. To know who one is requires to be oriented in moral space, a space in which questions can arise about what matters and what is good.

Idealism in the sense of actively committing oneself to ideals also plays a role in thinking about the formation of social policy and about politics in general. While the idealist will tend to seek social change in the form of reform or revolution, and therefore be progressive, the realist will tend toward more conservative policies. As in most idealisms (with the exception of the metaphysical idealism we find in Plato), the dignity of the human subject consists in its freedom.

Espen Hammer

See also A Priori and A Posteriori; Dialectic, in the History of Philosophy; Empiricism; Epistemology; Hegelianism and Contemporary Epistemology; Identity, Social; Language-Games and Forms of Life; Objectivity; Philosophy of History; Realism and Anti-Realism in the Social Sciences; Transcendental Arguments

Further Readings

Bradley, F. H. (1883/1978). *Appearance and reality: A metaphysical essay.* Oxford, England: Clarendon Press.

Dunham, J., Grant, I. H., & Watson, S. (2011). *Idealism: The history of a philosophy.* Montreal, Quebec, Canada: McGill-Queen's University Press.

Griffiths, M. (1992). *Realism, idealism, and international politics.* London, England: Routledge.

Hammer, E. (Ed.). (2007). *German idealism: Contemporary perspectives.* New York, NY: Routledge.

Marx, K. (1994). *Early political writings.* Cambridge, England: Cambridge University Press.

Offer, J. (2006). *An intellectual history of British social policy: Idealism versus non-idealism.* Bristol, England: Policy Press.

Pinkard, T. (2002). *German philosophy 1760–1860: The legacy of idealism.* Cambridge, England: Cambridge University Press.

Taylor, C. (1989). *Sources of the self: The making of the modern identity.* Cambridge, England: Cambridge University Press.
Weber, M. (2002). *The Protestant ethic and the spirit of capitalism.* London, England: Penguin.
Winch, P. (2007). *The idea of social science and its relation to philosophy.* London, England: Routledge.

Idealization in Social-Scientific Theories

Since the 17th-century scientific revolution, the role of theory in science has been the principal preoccupation of philosophy of science. The naive realist view, defended by some sophisticated realists, is that the central function of scientific theory is to explain, by reference to hidden unobservable mechanisms, the observable events of the real world. Idealizations within theory range from the use of abstraction to drastic simplifications of the subject matter being studied. Particularly with the rise of logical positivism, the realist view of theory was challenged by focusing on the role of idealization in scientific theory.

In this entry, the role of idealizations in the philosophy of the social sciences in general and the philosophy of economics in particular is examined.

Idealization and the Philosophy of Science

In Newtonian mechanics, for example, one represents physical bodies as point masses, even though it is self-evident that in reality the three-dimensional mass of a body is not located in a geometrical point. Since a premise known to be false cannot be used in any acceptable explanation, clearly scientific theories with idealizations known to be false fail to explain. An alternative account to that of naive realism is thereby required. This is furnished by instrumentalism. Instrumentalists, by focusing on the actual use of theories with idealizations, reject the realist notion of explanatory power. For instrumentalists, the realist question of the truth or falsehood of a theory is of no concern because a theory is basically a heuristic instrument or device for discovering more information about the world. Just as physical instruments/tools, such as a hammer, have specific, limited uses, a successful theory is a conceptual device used for the limited purpose of predicting or discovering new facts in a systematic way. Also, just as some physical instruments become obsolete and are replaced by more efficient ones, some theories become obsolete as heuristic-predictive instruments and are replaced by more efficient ones. Finally, just as physical instruments tell us little or nothing about the objects that result from their use, neither do scientific theories tell us anything about the observable world.

Idealization, Social Theorizing, and Economics

In the social sciences, neoclassical economics is frequently identified as the exemplar of successful theorizing. Within the domain of economics, especially for those committed to orthodox economics, there is no comparable successful rival: Neoclassical economics dominates the theoretical field. However, as the American economist Milton Friedman emphasized in the 1950s, neoclassical theory is akin to Newtonian mechanics, with its own specific range of idealizations. The idealizations embedded, for instance, in the notion of homo economicus—rational economic man—are evident in its assumptions, such as perfect foresight. Since these assumptions are patently false, any naive realist reading of neoclassical theory as a genuine explanation is without justification. What is required is an instrumentalist approach to the success of neoclassical economics. The criterion of success of neoclassical economic theory does not lie in the realism (truth or falsehood) of its assumptions; rather, it lies in whether or not neoclassical theory delivers correct predictions. According to Friedman, in this respect, neoclassical theory is eminently successful.

Friedman, however, gives a specific twist to this standard instrumentalist defense of the idealizations of neoclassical economics. This twist is related to the issue of the survival of firms in the marketplace. He maintains that, just as a billiard player will not survive in a billiard competition unless he behaves *as if* he were calculating his shots using the principles of Newtonian mechanics—which clearly he is not—a firm will not survive in the marketplace unless it behaves *as if* it were acting according to the principles of neoclassical theory—that is, profit maximization. Friedman's *as if* defense has been a major source of debate among both economists and philosophers of economics. The Cambridge

economist Frank Hahn rejects Friedmanite idealizations. For Hahn, Friedman's *as if* defense might be plausible if economic predictions attained the same level of success as quantum physics. Compared with quantum physics, however, neoclassical predictions are not impressive. Thus Hahn, contrary to Friedman and many others, insists that economics is *not* a predictive science á la physics. Rather, the aim of neoclassical theory is to gain an objective understanding of highly complex economic systems. To achieve this, one must avoid ambiguity, subjectivity, and vagueness, and hence the theoretical economist has recourse to logic and mathematics. This in turn means that idealizations based on the grounds of mathematical tractability are indispensable to economic theorizing. However, Friedmanite idealizations based on patent falsehoods are not acceptable because they hinder the fulfillment of the central task of gaining an objective understanding of an economy. A Hahn-like understanding, moreover, is quite different from hermeneutical approaches to understanding in the human sciences, emphasized in continental philosophy from Hans-Georg Gadamer to Paul Ricoeur.

Friedman's intriguing 1953 piece, "The Methodology of Positive Economics," remains a *locus classicus* for contemporary debates on idealizations in economics. Critical realists, like Tony Lawson, criticize Friedmanite idealizations because such assumptions hinder the discovery of the hidden generative mechanisms operating in real economies. On the other hand, realists, like Uskali Mäki, reconstruct scientific realism as minimal scientific realism and argue that in the end Friedman is a realist. Other equally interesting approaches have been developed. Alan Nelson, by recourse to a Cartesian model of idealization, as distinct from the standard idealization of physics, illuminates idealizations in economic theorizing.

Nelson attempts to retrieve a concept of idealization that he attributes to René Descartes, the 17th-century philosopher, mathematician, and scientist, which he terms *Cartesian idealization*. Descartes's ideas on idealization have received little attention, a neglect that is unwarranted for Nelson. He argues that Descartes's contribution on this topic could prove illuminating for economics, and by extension perhaps for social science in general. Cartesian idealization is embedded in a mechanistic view of science, whose laws are derived from fundamental metaphysical principles and are known to be true of the world, given their deduction from divinely underwritten innate ideas of God and matter. These ideal laws, which in a later terminology would be described as *a priori,* are true but for a variety of reasons as advanced by Descartes may be quite inadequate in providing accurate predictions of the behavior of physical objects. This was not a source of concern for Descartes, however, since he was satisfied that the provision of a possible mechanistic account based on unassailable metaphysical principles sufficed. In Nelson's interpretation of Descartes's conception of idealization, the Cartesian laws are true, even if seriously defective in providing accurate predictions, but they do provide schematic qualitative accounts of the mechanisms that may underlie the phenomena under study, which are not inconsistent with the ideal laws.

The task of understanding economics through Cartesian idealization is, for Nelson, neither disingenuous nor as outlandish as it might seem on first acquaintance. In fact, Nelson finds in the writings of Lionel Robbins, the distinguished economics methodologist, a position not incompatible with Nelson's own interpretation of Cartesian idealization. Robbins's methodological doctrine maintained that economics was a deductive system, based on a set of assumptions that are so integral to everyday experience that, once stated, they are deemed self-evidently obvious. If these assumptions relate to reality, as Robbins accepted they did, then the deductions had a similar point of reference in reality. However, in economics, there are acute problems arising from whether the deductions from the assumptions refer to reality, given the abysmal performance of economics in quantitative prediction. This particular concern has been at the center of a major methodological debate that dominated a large part of the 20th century. Milton Friedman reoriented this question by insisting that predictive accuracy was the central task of economic theory, as we saw above, and that the theory, complete with its set of assumptions, that produced the best set of predictions was the best theory; and its assumptions were then retrospectively or instrumentally justified, regardless of whether they were "realistic" or not, where "realistic" has the connotation of empirical accuracy. Consequently, economists committed to Friedman's instrumentalism and invoking standard idealization, as interpreted by Nelson, would argue

that the assumptions of economic theory are false, while those advocating Cartesian idealization, as interpreted by Nelson, would require the assumptions of economic theory to be true even if neither realistic nor empirically accurate. Nelson provides an interesting contrast between the concepts of standard idealization, derived from physics and adhered to by numerous economists, and Cartesian idealization, which provides an alternative view of idealization that arguably requires more reflection than it has received to date.

Rationality and the Social Sciences

Homo economicus of neoclassical economics is an idealized account of rationality in the pursuit of wealth. This idealization has been increasingly extended to the other sociohistorical sciences. According to Popperians, the aim of the social sciences is to understand/explain social events, and to this end, social scientists construct models. These models are idealizations in the sense that a model reconstructs the actual situation in terms of a *typical* social situation. These idealized models on their own, however, are not sufficient to explain social events: Recourse to the rationality principle is also required. In this connection, Popperians use the homo economicus idealization as a benchmark against which the rationality of a human action is measured. This homo economicus idealization subsists in Karl Popper's autonomous, quasi-Platonistic World 3. In this Popperian approach, the homo economicus model is not part of the description of social actions. Rather, it is the prescriptive benchmark for deciding the rationality, the degree of rationality, or the irrationality of actual actions. Thus, homo economicus becomes the logic of rational choice.

According to other philosophers, however, it is very evident that the idealized model of homo economicus does not apply across the broad range of human interactions. Rather, as Philip Pettit suggests, people's actions are culturally framed in a vast network connecting their commitments to ideals, fair play, friendship, honesty, integrity, kindness, membership of certain groups, politeness, and so on. Rationality in this extensive and thick cultural setting is very far from the model of the self-interested homo economicus. Pettit, however, argues that the acknowledgment of the cultural framing of rationality does not imply the rejection of the homo economicus model. He proposes a conciliationist thesis in which both are reconcilable. Clearly, while the principles of homo economicus are not actually present in culturally framed decisions and choices, they are virtually present—they hover on the edge of realization. This virtual presence becomes evident when an agent's decisions made in his or her cultural frame are liable to cost him or her dearly in self-regarding terms. In such situations, the alarm bells of homo economicus begin to ring, and the agent begins to calculate à la homo economicus.

Clearly, idealizations in both economics and the other sociohistorical sciences raise intriguing methodological and philosophical issues and debates, which continue to exercise philosophers of the social sciences.

Thomas A. Boylan and Paschal O'Gorman

See also Critical Realism in Economics; Homo Economicus; Instrumentalism of Scientific Theories and Constructive Empiricism; Models in Social Science; Philosophy of Economics, History of; Rationality and Social Explanation; Realism and Anti-Realism in the Social Sciences; Social Choice Theory

Further Readings

Brzeziski, J., Coniglione, F., Kuipers, T. A. F., & Nowak, L. (Eds.). (1990). *Idealization I: General problems.* Amsterdam, Netherlands: Rodopi Press.

Elgin, C. (2009). Exemplification, idealization, and scientific understanding. In M. Suarez (Ed.), *Fictions in science: Philosophical essays on modeling and idealization* (pp. 77–90). London, England: Routledge.

Friedman, M. (1953). *Essays in positive economics.* Chicago, IL: University of Chicago Press.

Godfrey-Smith, P. (2009). Models and fictions in science. *Philosophical Studies, 143,* 101–116.

Hahn, F. H. (1984). *Equilibrium and macroeconomics.* Oxford, England: Blackwell.

Hamminga, B., & De Marchi, N. B. (Eds.). (1994). *Idealization VI: Idealization in economics.* Amsterdam, Netherlands: Rodopi Press.

Lawson, T. (2003). *Reorienting economics: Economics as social theory.* London, England: Routledge.

Mäki, U. (Ed.). (2002). *Fact and fiction in economics: Models, realism and social construction.* Cambridge, England: Cambridge University Press.

Nelson, A. (2001). Two models of idealization in economics. In U. Mäki (Ed.), *The economic world view: Studies in the ontology of economics* (pp. 359–368). Cambridge, England: Cambridge University Press.

Pettit, P. (2002). *Rules, reasons and norms.* Oxford, England: Clarendon Press.
Robbins, L. (1945). *An essay on the nature and significance of economic science.* London, England: Macmillan.
Suarez, M. (Ed.). (2009). *Fictions in science: Philosophical essays on modeling and idealization.* London, England: Routledge.

Identity, Personal (Philosophy of)

The entry explains what the philosophical problem of personal identity is, presents the main solutions offered as well as some recent developments, and highlights the relation between philosophy and the social sciences with respect to this problem.

The philosophical problem of *personal identity* is most commonly understood as a special case of a more general metaphysical puzzle about the persistence of complex objects. The general puzzle is expressed in the famous case of the Ship of Theseus, whose planks are replaced one by one over several years until the ship that remains has no wood in common with the original. The question is *whether a single object persists throughout these changes* and, if so, in virtue of what it remains selfsame despite the replacement of all of its parts. The problem of personal identity applies this question to persons. Persons too change drastically over time, and philosophers of personal identity seek a systematic account of the kinds of vicissitudes they can and cannot survive.

The standard philosophical question of personal identity is thus different from the question of personal identity as it is construed in most other contexts. Often, worries about identity involve questions about someone's fundamental psychological or social makeup rather than what makes her a single persisting entity. An adolescent trying to find her identity, for instance, is not uncertain about which individual object she is but about what she truly believes and desires. There are, however, important points of contact between the two questions, and understanding this overlap leads to many fruitful interactions between philosophy and the social sciences.

Psychological Continuity

One powerful strand of thought in the philosophy of personal identity, found, for instance, in the work of Derek Parfit, John Perry, Sydney Shoemaker, and David Lewis, is the view that personal identity over time should be defined in terms of *psychological continuity.* This approach is an updated version of John Locke's claim that the identity of persons resides in "sameness of consciousness" rather than sameness of substance. To defend this view, Locke asks us to imagine the consciousness of a prince entering the body of a cobbler and replacing the cobbler's own. The result is someone with the cobbler's physical appearance but with the memories, intentions, beliefs, and desires of the prince. Clearly, Locke says, this individual is the prince and not the cobbler, which shows that the *person* goes where the psychological life goes. Present-day psychological theorists take up and develop this general insight, broadening Locke's notion of sameness of consciousness into a more general notion of psychological continuity and defending the suggestion that it can define the literal persistence of persons. This approach does not claim that the metaphysical identity question raised by philosophers is the same as the more common psycho/social identity question, but it does see the two as importantly related.

A central task for psychological theorists is to provide a viable understanding of what is involved in identity-constituting psychological continuity. One element usually taken to be an important part of such continuity is *memory.* Locke is typically taken to hold that what makes a person at one time a person at an earlier time is that the later person has an autobiographical memory of an experience of the earlier person's experiences. There are several difficulties with this position as stated, and some question about why only memory should count as an identity-constituting connection, but psychological continuity theorists generally agree that memory is an important part of what constitutes personal identity. To understand this claim and make it plausible, it is necessary to explore a variety of questions about the nature of memory—what kinds of memory there are, how different sorts of memory connect to our sense of self and psychological functioning, what it means for a memory to be accurate, and the extent to which memory is possible without other kinds of psychological continuity being in place. All these questions require close attention to empirical work.

Some psychological theorists suggest that the best way to think about identity-constituting psychological continuity is in terms of a *self-narrative*. The distinctive psychological lives of persons, these theorists argue, take the form of a narrative, and it is because we have narrative self-understanding that we are able to have the kinds of experiences and take the kinds of actions that make us persons. The unity of an individual person, on this view, is the unity of a *life-narrative*.

The narrative conception of persons has also been an object of investigation in psychology, for instance, in the work of Jerome Bruner, Ulrich Neisser, and Katherine Nelson. Nelson's work emphasizes the development of narrative self-understanding in children and its connection to the emergence of the peculiar kind of self-consciousness humans possess. This kind of self-awareness is very like the consciousness in terms of which Locke defines personhood and personal identity, and a more complete understanding of its origins and structures is invaluable in defining the kind of narrative unity that potentially constitutes the metaphysical unity of persons.

The Role of Social Context

Another central set of issues in the philosophy of personal identity involves the question as to what extent and in what ways social environments and interactions can play a role in constituting the metaphysical identity of a person. Some philosophers hold that relational properties can in no way define the true identity of any entity, since the entity must first exist as a well-defined whole before it can stand in a relation to anything else. Others, however, insist that a person as subject and agent can exist only in a social context within which she or he relates to and distinguishes herself or himself from others. In exploring this question, the burgeoning and wide-ranging body of empirical work on the ways in which social embeddedness scaffolds cognitive development and affects self-understanding and self-consciousness offers important insights. It is only with a detailed picture of the various ways in which social relations affect our inner lives that we are in a position to assess whether these might be constitutive of metaphysical identity.

Unity of Consciousness

Finally, there are questions about whether the unity of consciousness in terms of which Lockeans define the unity of a person is a real phenomenon. Philosophers like Daniel Dennett have used findings from psychology to argue that the unified subject is an illusion and should be considered a useful fiction rather than a metaphysical fact in need of explanation.

Philosophy and Social Science

Philosophers of personal identity seek to answer a metaphysical question about the persistence of an object. Many wish to couch this answer in social or psychological terms, and this means appreciating and respecting empirical findings from the social sciences. At the same time, philosophical distinctions among different kinds of identity questions and their connections to one another provide a framework that is helpful in interpreting empirical results and understanding their broader significance. Philosophy and the social sciences thus have a great deal to offer one another in exploring these issues.

Marya Schechtman

See also Consciousness; Death and Immortality, Philosophical Perspectives; Personal Identity and Trauma; Self and Essential Indexicality; Self-Knowledge

Further Readings

Bruner, J. (1985). *Actual minds, possible worlds.* Cambridge, MA: Harvard University Press.

Locke, J. (1975). *An essay concerning human understanding* (P. Nidditch, Ed.). Oxford, England: Clarendon.

Martin, R., & Barresi, J. (Eds.). (2003). *Personal identity.* Malden, MA: Blackwell.

Parfit, D. (1987). *Reasons and persons.* Oxford, England: Oxford University Press.

Paul, E. P., Miller, F. D., & Paul, J. (Eds.). (2005). *Personal identity.* Cambridge, England: Cambridge University Press.

Schechtman, M. (1996). *The constitution of selves.* Ithaca, NY: Cornell University Press.

Identity, Social

According to identity theory, identities define who a person is in terms of the groups or categories to which they belong (*social identities*), the roles they occupy (*role identities*), and the personal characteristics they claim (*person identities*). For example,

an individual's social identity as an American or a Hispanic is what it *means* to him or her to be an American or a Hispanic, an individual's role identity as a truck driver or a student is what it *means* to that individual to be a truck driver or a student, and an individual's person identity as a dominant person or a moral person is what it *means* to that individual to be dominant or moral. The focus on *meaning* comes from identity theory's origins in *symbolic interactionism*.

"Social," "person," and "role" are three different bases for identities, though each functions in the same way. This entry will discuss these identity bases in turn after having first outlined the identity components and the so-called identity verification process whereby identities are maintained and protected in social contexts. Such a verification process is further explained, and its pivotal role in building these identity bases in social contexts is stressed.

Introduction

The *meanings* of identities are derived from the culture that is shared among persons so that the understanding of the meanings is shared and communicated. People act to create and maintain meanings in interaction with others that identify and reflect who they are. They do this through their words, actions, dress, and appearance. Truck drivers act and dress like truck drivers. They have and maintain the tools, equipment, licenses, skills, and language that are necessary to be truck drivers. In this way, they identify themselves as truck drivers, and others can identify them as truck drivers and interact appropriately with them, whether those others are other truck drivers, dispatchers, truck-stop personnel, or loaders. In addition to *displaying* their identities, people also *protect* and *maintain* their identities (this is the "identity verification process") when the meanings in the situation are changed by circumstances or by others verifying their own identities.

Identity Components

To understand the identity-verification process, let us begin with an outline of the component parts of an identity, which are the same for role, person, and social identities.

The first part is the *identity standard,* which is the set of meanings defining the identity. These are the meanings that the person displays and maintains while in the identity. Second are the meanings of the *output* or *behaviors* enacted in the situation. In identity theory, it is not the behavior as such that matters but how that behavior is interpreted—what it means. People act to reproduce in the situation the meanings held in the identity standard. The third part is the *perceptual input* of identity-relevant meanings from the situation. People perceive how they are coming across to others, either directly (direct appraisals) or indirectly through what are called *reflected appraisals*. Reflected appraisals are what each person thinks others think of him or her. For example, if I act in such a way as to make myself understood as a dominant person but the reaction of others indicates that they think I am meek (i.e., I think they see me as meek—the reflected appraisal), then I am not getting my message across, and I will act in an even more dominant manner.

The fourth part is the *comparator,* which functions to compare the perceived meanings with the identity standard meanings. This comparison is a simple difference. When the difference is zero, the person perceives that the meanings she is portraying in the situation match those held in her identity standard. When they match, her identity is being verified. Others see her the way she sees herself. According to identity theory, people control their perceptions of identity-relevant meanings to have them match the meanings in their identity standard. They allow their behavior to vary in order to keep the perceived meanings constant, matching the identity standard that defines them.

Verification Process

In identity theory, people are motivated to verify their identities—to make sure the meanings in the situation reflect the meanings held in the identity standard. One could view this as a goal, to make the situational meanings match the identity standard meanings. This is true for all identity bases: social identities based on group or category memberships, role identities based on the roles one occupies, and person identities that make one a unique individual. When verification does not occur for any of these identities, when the reflected appraisals do not match the identity standard, there are several outcomes. First, there is a negative emotional response to nonverification. The strength of the negative emotion is a function of the squared amount of difference between the reflected appraisal meanings and the identity standard meanings. When a person has a certain level of dominance as an important meaning

in their person identity standard, she becomes upset if others see her either as more dominant or as less dominant than defined in the identity standard.

At the same time a person is feeling an emotional response, she also begins to act in ways that change the displayed meanings in the interaction situation. The behavior is a function of the plain (not squared) difference between the reflected appraisal meanings and the identity standard meanings. If a person is seen, for example, as less dominant than the identity standard, that person will increase the dominance meanings in her behavior. The person may act gruffer, more abrupt, more commanding, more forceful, or some combination of these or other behaviors that convey the appropriate meanings to impress upon others who she is. By perceiving others' reactions that convey impressions, the reflected appraisals come to match the identity standard. Similarly, if the person is seen as aggressive and authoritarian, that impression is more dominant than what is defined in the identity standard, and the person will act in ways that convey less dominance until the reflected appraisals are brought into agreement with the identity standard.

In addition to the immediate emotional response and the adjusted behavior to achieve identity verification, there is a longer-term consequence of maintaining or not maintaining verification over time. This longer-term effect influences the self-esteem of the person. A person who continues to have difficulty verifying an identity, especially an important identity, will lose self-esteem. Being able to achieve verification of important identities over time will increase feelings of self-esteem. In this way, self-esteem is earned or lost by one's achievements with respect to identity verification. Self-esteem cannot be given to a person; rather it is earned by a person in his own mind by his own successes or failures at verification. Identity theory sees self-esteem as made up of three parts or bases: self-worth or feeling of value, self-efficacy or feeling competent and able to accomplish things, and authenticity or feeling that you can be your true self.

Identity Bases

There are three different bases for the identities that persons have. *Social* identities are based on group or category membership, *role* identities are based on the roles a person occupies within a group or organization, and *person* identities denote the meanings each person claims as defining and describing him- or herself as a distinct person. While identities from each basis have the same component parts and each operates in the same manner for verification, they differ in the nature of the meanings they contain and the consequences of verification or nonverification.

Social Identities

The meanings in a social identity define a person to be *similar to* others in their group or category and to be *different from* persons in other contrasting groups or categories. All the persons in a fraternity, for example, define themselves in terms of upholding the principles of the fraternity and being different from other fraternities, perhaps emphasizing academics over sports. By verifying this social identity, persons help maintain the principles and distinctions of the fraternity, and they maintain the distinctions and separations from other fraternities. By being like others, they are accepted by others in the fraternity for who they are; they belong and their self-worth, one of the bases of self-esteem, is maintained or enhanced.

Role Identities

The meanings in a role identity do not make persons like others with whom they interact (as in the case of social identities), but they make them *complementary* to role partners, that is, persons occupying counter-roles. The role of a student is complementary to the counter-role of a professor. The role of a daughter is complementary to that of the mother. The role of fraternity president is complementary to the counter-roles of the other officers in the fraternity. Verifying a role identity means carrying out the expectations and standards of the role, and this in turn allows the occupants of counter-roles to carry out their duties and obligations—that is, verify their role identities. Verification of a role identity is an accomplishment, and as a consequence, it leads to an increase or maintenance of feelings of self-efficacy or competence, the second basis of self-esteem.

Person Identities

The meanings in a person identity are those that make that person a *unique* individual. These meanings, like dominance or sincerity or morality, are always with the person, in all situations, in all groups, and in all roles. In that way, these meanings

are very central or core to the person, and verification of these meanings allows the person to be who he or she is truly as an individual. Verification of a person identity leads to an increase or maintenance of feelings of authenticity, the third basis of self-esteem.

Mutual Verification

Because identity verification takes effort to maintain meanings in the face of disturbances or events in the situation that change meanings away from the identity standard, it is advantageous to interact with others who, in being themselves, allow us to be ourselves. When husbands and wives, for example, verify their partner's identity in the process of verifying their own identity, we have a situation of mutual verification; they are supporting each other's identity meanings. Mutual verification has been found to raise the self-esteem of the partners, increase their love and trust, and strengthen their bonds as a couple or group. They become a "we" rather than two "I"s. Mutual nonverification, on the other hand, has been found to lead to separation and divorce.

Identity and Society

Because society is made up of groups and organizations with their roles and divisions of labor, as well as categories of persons who may or may not be granted access to these groups, roles, and organizations, it is clear that identity verification by maintaining the categories, groups, roles, and organizations also maintains society. This is the link between individuals and society. Individuals hold identities that link them to positions in society (roles and groups) that are defined by the shared culture. Because these identities are self-defining, they are enacted and protected. It is who we are. But by enacting and protecting identities, people enact the roles, maintain the groups, and keep the divisions and separations between social categories. It is clearly a circular process. Societies and cultures define the roles, groups, and categories. Individuals take these roles, groups, categories, and organizations on as identities and reproduce the meanings that define themselves, their roles, their groups, and their organizations, that is, their society. In the process, they feel good and gain self-esteem.

However, this is not always possible. Sometimes (perhaps often) people have trouble verifying their identities because others who are more powerful maintain meanings in situations that do not fit their identity standard. This, of course, leads to negative emotions and, over time, an erosion in self-esteem. Conflict can arise as people compete for particular meanings in the interaction that fit with their identities. People may leave situations and others who do not verify their identities, they may abandon identities that are not verified, and identities may change over time so that they fit the meanings being provided by more powerful persons. Indeed, according to identity theory, all of these things happen. Identity theory also tells us some of the conditions under which these happen, alone or in combination.

Peter J. Burke

See also Emotions; Group Identity; Personal Identity and Trauma; Self and the Social Sciences; Symbolic Interactionism

Further Readings

Burke, P. J., & Harrod, M. M. (2005). Too much of a good thing? *Social Psychology Quarterly, 68,* 359–374.

Burke, P. J., & Reitzes, D. C. (1981). The link between identity and role performance. *Social Psychology Quarterly, 44*(2), 83–92.

Burke, P. J., & Stets, J. E. (2009). *Identity theory.* New York, NY: Oxford University Press.

Stets, J. E., & Burke, P. J. (2000). Identity theory and social identity theory. *Social Psychology Quarterly, 63*(3), 224–237.

Stets, J. E., & Burke, P. J. (2003). A sociological approach to self and identity. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 128–152). New York, NY: Guilford Press.

Stryker, S., & Burke, P. J. (2000). The past, present, and future of an identity theory. In The state of sociological social psychology [Special issue]. *Social Psychology Quarterly, 63*(4), 284–297.

IDEOLOGY

Ideology is one of the—if not *the*—most complex concepts in the social sciences. It is what has been called an "essentially contested" concept; that is, one whose definition (and therefore application) is a subject of acute and unavoidable controversy.

This is not surprising since the concept of ideology deals with the bases, if any, of our most fundamental values. This entry gives an account of the origins of the term, goes on to introduce the main Marxist conception of ideology, and concludes with an account of recent and current developments.

Origins

The term *ideology,* unlike many other similarly controversial terms, such as *democracy* or *liberty,* is relatively new. It was coined in the immediate aftermath of the French Revolution by the French philosopher Destutt de Tracy (member of the "Société des idéologues") in his *Elements d'Ideologie,* written between the years 1801 and 1815. Here, he proposed a new *science of ideas* and *idealogy,* which would be the ground of all other sciences. Rejecting the concept of innate ideas, de Tracy explained how all our ideas are based on physical sensations (and hence ideology is a branch of zoology). A rational investigation of the origin of these ideas, free from religious or metaphysical prejudice, would be the foundation of a just and happy society. For the investigation of individual ideas would show their common origin in universal needs and desires. These needs would form the framework of laws regulating society on a natural basis and promoting the harmonious fulfillment of the relevant desires. For the natural and the social worlds coincided. And this coincidence would be laid bare by the rational assessment of the origin of ideas, by ideology. Clearly, here the notion of ideology was positive and progressive (vide Napoleon Bonaparte's reaction to it).

In the German tradition, by contrast, the Romantic movement laid emphasis on the way we invest the world with our own meanings. Unlike de Tracy, for whom the natural and the social worlds were pellucid to the rational mind, the German romantics considered that human beings collectively and individually created their own reality in response to changing circumstances. For the German Idealist philosopher G. W. F. Hegel, who tried to give these ideas a systematic intellectual basis, the ideas of a particular age (*Zeitgeist*) could not claim absolute validity in themselves for they were evidently relative to changing historical situations. If there *were* a rationality, a meaning to history, it would have to be found in the whole process rather than in the partial aims of particular individuals and epochs.

Marx and His Followers

These Hegelian ideas strongly influenced Karl Marx, and it was Marxism, mixing both the French and the German trends, that put the concept of ideology in the forefront of political discourse. Marx himself did not (most of the time) think that all ideas were ideology and did not wish simply to produce a more dynamic version of de Tracy's science of ideas. For him, it was their connection with the class struggle and its social and economic basis that gave certain ideas their ideological force. Society was in fact driven by conflicts of interest, but for it not to fall apart, these oppositions were *covered up* by ideas that represented attempts to portray society as cohesive rather than conflictual by justifying the asymmetrical distribution of social and economic power (i.e., an "ideological distortion"). Hence, Marx's famous statement that "the ideas of the ruling class are in every epoch the ruling ideas," since those who controlled economic production and distribution could also regulate the production and distribution of ideas. In his late writings, Marx added a further dimension. Here, ideology still served, as in the earlier conception, to conceal and invert real relationships between people. But in addition, Marx now concentrated on the way in which ideology was produced by the form in which people related together in the ordinary everyday transactions of capitalist society. There was also less emphasis on illusion, since ideology was seen here as reflecting something real, if decidedly partial, and also as being itself a real force.

During the first half of the 20th century, particularly under the influence of Marx's close collaborator, Friedrich Engels, the positivist tradition of de Tracy, with its attempt at an objective science of society, was prominent. Lenin was happy to talk of "socialist ideology," with no pejorative connotation. Perhaps the subtlest Marxist thinker was the Italian Antonio Gramsci, who drew a close parallel between ideology and his concept of *hegemony,* which he inherited from Lenin and Lukacs. For Gramsci, the worldview of the ruling class was so thoroughly diffused by its intellectuals as to become the "common sense" of the whole of society, the "structure of feeling" in which it lived. This conception was Gramsci's answer to the puzzle of how capitalism had managed to survive in the bourgeois democracies of the West.

Outside the Marxist tradition, Marx's two fellow-founders of sociology, Max Weber and Émile Durkheim, produced discussions about the genesis

and validity of ideas that contributed substantially to subsequent treatments of ideology—Weber as the inspirer of much of the later empirical investigation in the Anglo-Saxon world and Durkheim as the forerunner of structuralist analyses. But it was the German sociologist Karl Mannheim who produced a comprehensive theory of ideology that is still a reference point for today's discussions. Mannheim, in his 1929 book *Ideology and Utopia,* evolved a sociology of knowledge that investigated, through description and structural analysis, the ways in which social relationships influence thought. Avoiding relativism, Mannheim attached a privileged role to intellectuals due to their constituting a relatively classless stratum that was not too firmly situated in the social order.

The End of Ideology?

In the last half of the 20th century, discussion has been dominated by the "end of ideology" thesis. This positivist and restrictive definition of ideology, combined with a discussion of ideology that, itself, claimed not to be ideological, flourished mainly in the United States. This reflected the predominance of behavioralism in postwar American political science and the stress on ordinary language and linguistic analysis in British philosophy. Its main proponents were Seymour Martin Lipset, in his influential 1960 book *Political Man,* and Daniel Bell, whose *End of Ideology,* published the same year, claimed that ideology, which was once a road to action, has become a dead end. Their approach was supported by Hannah Arendt's *Origins of Totalitarianism,* which saw all ideological thinking as suffused with totalitarian elements.

More recently, some have seen postmodernism as a recycling of the end-of-ideology thesis: Critical and mockingly destructive, it rejects even broad and tentative attempts to establish foundations for our thought, foundations that are swept away in a spiral of endless metaphor. An accessible form of this approach is found in the work of the contemporary American philosopher Richard Rorty, whose playful and "ironic" liberalism (as he called it) springs from a postmodernism that expresses the continental tradition of Friedrich Nietzsche and Martin Heidegger through an analytical Anglo-Saxon pragmatism. The most well-known proponent of the end of ideology, and indeed of history itself, is Francis Fukuyama, whose book *The End of History and the Last Man* elaborated many of its themes.

The beginning of the 21st century, however, has shown these views to be premature. The resurgence of religion and the challenges provided by environmental change show that ideology, under almost any of its manifold definitions, is with us for a good time to come. And it is difficult to avoid the useful conclusion that all views about ideology are themselves in some sense ideological. But if we are indeed all implicated in ideologies that are both real and powerful, an understanding of this fact can at least prevent us from becoming their unconscious victims.

David McLellan

See also Alienation: From Philosophy to Sociology; Capitalism; Dialectic, in the Social Sciences; Enlightenment, Critique of; Marxism and Social/Historical Explanation; Modernity

Further Readings

Bell, D. (2000). *On the exhaustion of political ideas in the fifties.* Cambridge, MA: Harvard University Press.
Eagleton, T. (2007). *Ideology: An introduction.* London, England: Verso.
Heywood, A. (1998). *Political ideologies: An introduction.* Basingstoke, England: Macmillan.
Larrain, J. (1979). *The concept of ideology.* London, England: Hutchinson.
Lichtheim, G. (1967). *The concept of ideology.* New York, NY: Random House.
McLellan, D. (1995). *Ideology.* Buckingham, England: Open University Press.
Seliger, M. (1976). *Ideology and politics.* London, England: Allen & Unwin.
Therborn, G. (1999). *The ideology of power and the power of ideology.* London, England: Verso.

Implicit Bias and Social Cognition

Recent decades have seen large advances in the understanding of social cognition. One set of unsettling findings is about what have come to be called *implicit biases,* unconscious negative evaluative tendencies about individuals based on their membership in a social group. The existence and character of implicit biases raise a variety of interesting

issues for different areas of philosophy, including many relevant to the social sciences. Getting into a position to address those issues requires becoming familiar with implicit biases themselves, the sorts of tools psychologists have developed to investigate them, and some of their more noteworthy features. This entry addresses the aforementioned issues first. The final section briefly describes some of the initial implications of implicit biases that have been drawn for philosophy and the social sciences.

Useful Distinctions

Implicit biases may not fit easily with an intuitive picture of the mind, so social psychologists have made a number of clarifying distinctions to characterize them. A mental state or process is *automatic* if it is typically outside a person's conscious control and can influence a person's behavior and judgment without deliberation, attention, or effort. Many, though not all, automatic mental states or processes are also *implicit*; they are outside the person's conscious awareness. To study automatic and implicit mental states, social psychologists have developed a number of sophisticated experimental techniques, often called indirect measures. Much of what is known about implicit biases has been inferred from people's performance on tasks that use such techniques. The measures are said to be *indirect* in that they do not directly rely on subjects' powers of introspection or self-report and are able to avoid the problems associated with both. In these tasks, people who harbor implicit biases show a slight but consistent negative evaluation of members of whatever social group they are implicitly biased against, relative to members of other social groups. For instance, those with implicit biases against the elderly will be more likely to show a slight preference for young and middle-aged people, those with implicit biases against women will show a preference for men, those with implicit biases against Blacks will show a preference for Whites or Latinos, and so on. Implicit biases directed toward social groups defined along a number of dimensions have been found, including race, gender, sexual orientation, age, weight, and religion. Details about implicit biases (the social groups they are directed at, their strength, and even their presence) can vary from person to person as well as from culture to culture (though little cross-cultural research has been done).

Features of Implicit Social Cognition

One consistent and striking finding of this research is that within a single person, implicit biases against a particular social group are distinct from, and can coexist with, consciously endorsed attitudes to the contrary. For example, when directly asked, a person might sincerely state about herself that she is not racist and holds tolerant, egalitarian views toward all races, thus expressing her *explicit* attitudes. Nevertheless, indirect testing can reveal that she also harbors *implicit* biases against members of certain races. Indeed, research has shown that it is possible to harbor implicit biases against social groups that one is a member of oneself—that is, one's own race, gender, and so on.

Initial studies into development suggest that though implicit biases are not easy to uproot, they are fairly easy to acquire. Certain types, implicit racial biases, appear to be widespread, at least in the United States. The influence of implicit biases on behavior, especially behavior in the real world, outside controlled laboratory experiments, is difficult to measure with precision, but mounting evidence indicates that implicit biases toward a particular social group can subtly shape a person's evaluations of and interactions with members of the social group to which they apply.

Another area of research explores the ways in which implicit biases might be managed or controlled. For instance, once a person becomes aware that she harbors implicit biases toward a certain social group, she might make an effort to explicitly suppress the expression of those biases, in both judgment and behavior. This has been shown to work to some extent but requires vigilance, is mentally fatiguing, and can backfire in a number of ways. Other forms of self-control not based on suppression, such as making an active attempt to see others as individuals rather than as members of a category or social group, appear to fare better. Moreover, taking less direct steps, such as exposing oneself to positive images of, or having positive interactions with, members of the social group in question, has also been shown to lessen the influence of implicit biases.

Philosophical Issues

Philosophers have just begun to take note of implicit biases and to examine their implications. The work of those who have can be divided into two broad groups. First, those interested in the epistemology

and conceptual foundations of the social sciences, especially those who think that many of the central concepts (races, genders, classes, etc.) pick out kinds that are *socially constructed,* have begun exploring how features of social cognition affect the construction and stability of those concepts and kinds. They have emphasized how full explanations of the sorts of regularities studied by social scientists can incorporate, and sometimes may require, appeal to features of individual psychologies, including elements of social cognition like implicit biases.

Second, those interested in social justice have seen implicit biases as relevant to a number of issues. Some have attempted to understand how they contribute to social inequalities like persistent gender and racial disparities, emphasizing that the moral problems raised by such cases are intertwined with the epistemic burdens imposed by the implicit biases involved. Philosophers are also beginning to investigate how social and legal institutions might best take implicit biases into account and assign responsibility for individual behaviors and institutional outcomes that are influenced by them. Another preliminary line of inquiry has attempted to evaluate various proposals for social reform in light of what is known about social cognition and implicit biases.

In sum, research on implicit biases, into the psychology of the biases themselves, their influence on the behavior of institutions, and their implications for understanding groups and population-level social dynamics, is barely out of its infancy; there remains much fertile and exciting territory to explore.

Daniel Kelly

See also Prejudice and Stereotyping; Race, Theories of; Racial Critiques of Social Science Applications; Social Cognition; Social Constructivism; Unconscious Social Behavior

Further Readings

Gendler Szabó, T. (2011). On the epistemic costs of implicit bias. *Philosophical Studies, 156,* 33–63.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*(1), 17–41.

Kelly, D., Faucher, L., & Machery, E. (2010). Getting rid of racism: Assessing three proposals in light of psychological evidence. *Journal of Social Philosophy, 41*(3), 293–322.

Kelly, D., Machery, E., & Mallon, R. (2010). Race and racial cognition. In J. Doris (Ed., with The Moral Psychology Research Group), *The Oxford handbook of moral psychology* (pp. 433–472). New York, NY: Oxford University Press.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test: IV: What we know (so far) about the method. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies* (pp. 59–102). New York, NY: Guilford Press.

Machery, E., Faucher, L., & Kelly, D. (2010). On the alleged inadequacies of psychological explanations of racism. *The Monist, 93*(2), 228–255.

Mallon, R., & Kelly, D. (2012). Making race out of nothing: Psychologically constrained social roles. In H. Kincaid (Ed.), *The Oxford handbook of the philosophy of the social sciences* (pp. 507–532). New York, NY: Oxford University Press.

Saul, J. (in press). Unconscious influences and women in philosophy. In F. Jenkins & K. Hutchison (Ed.), *Women in philosophy: What needs to change?* Newcastle upon Tyne, England: Cambridge Scholarly Press.

Individualism, Methodological

Methodological individualism is a principle, or rule, for the analysis of social phenomena, which says that they should invariably be understood and explained in terms of individual human beings: their motives, actions, interaction, and situation. At least, this is the main version of this much-debated doctrine. As this entry will show, methodological individualism is not one but many, and its various versions are united by family resemblance rather than by the invariable common features found in all versions. What they do share in common is a focus on individuals. They all assign pride of place to individuals in their approach to society.

Methodological individualism originates with the Austrian School of Economics at the end of the 19th century. The height of the debate about it was in the middle of the 20th century, when the philosopher Karl Popper and his followers made it a central part of their methodology.

Methodological individualism was advanced very much against its opposite: *methodological holism,* or *collectivism,* which says that social phenomena should be understood and explained in terms of social wholes and/or collectives, such as social structures and social institutions.

Among the social sciences, economics is, by far, the most and anthropology probably the least individualistic in the methodological sense of the term. Political science and sociology fall somewhere in between. The main reason for this is that methodological individualism is part of rational choice theorizing, which is the approach of mainstream economics. The prevalence of methodological individualism in the other social sciences is largely a function of the extent to which they use rational choice.

Methodological Individualism *Avant la Lettre*

Methodological individualism was used long before it was explicitly stated. Already in Greek antiquity, we find it in the theories of society suggested by the Sophists and by Epicurus. This is the archetype of methodological individualism, and it reappeared in the modern theory of the social contract, which dominated thinking about society from the Renaissance to the 18th century. Most well-known are the theories of Thomas Hobbes and John Locke. The former, in particular, is recognized as the prototypical methodological individualist *avant la lettre*. He explained social institutions as the intended results of the actions of rational and self-interested individuals in a state of nature.

A second paradigm of methodological individualism appeared with the rise of economics. Adam Smith analyzed market phenomena as the unintended consequences of the actions of individuals, and this analysis eventually developed into the theory of general equilibrium. The aim of this theory is to do away with all social institutions and relations in models of the market. According to its founder, the French mathematical economist Léon Walras, pure economics is a natural science of things, recognizing only isolated individuals and impersonal prices.

Of the classical economists, only John Stuart Mill was much interested in methodology, and his methodology was individualistic. In *A System of Logic* (1843), he argued that all social laws are based on psychological laws of the mind. Only psychological laws are causal. Social laws, therefore, can and should be reduced to psychological laws. This methodological individualist view was adopted by neoclassical economics, in the form of a belief that macroeconomics should be provided with microfoundations.

Austrian Methodological Individualism

As an explicitly stated principle of social analysis, methodological individualism probably goes back to the Austrian economist Carl Menger (*Problems of Economics and Sociology,* 1883). He suggested that complex phenomena should be reduced to their simplest element, the actions of individual human beings, and called it "the atomistic method." Complex social phenomena were analyzed, by Menger, as the unintended consequences of the actions of individuals. The term *methodological individualism* was coined by the eminent economist and political scientist Joseph Schumpeter, who used it to make a distinction between methodological and political individualism.

Max Weber brought methodological individualism from economics to sociology. According to him, sociology too should use a strictly individualist method and treat the single individual as its basic unit, or atom. Weber's individualism follows from his interpretive sociology, which takes its point of departure in the interpretive understanding of social action. Understanding is achieved when we know the subjective meaning individuals attach to their actions. This includes an understanding of their motives, which are the immediate causes of their actions and the ultimate causes of all social phenomena.

Weber's interpretive sociology exerted a strong influence on the Austrian economist Ludwig von Mises, who turned it into a nominalist and subjectivist ontology of the social world (nominalist in the sense that it does not assign ontological status to social wholes, it does not admit them as irreducibly real). According to him, only individuals exist, and society, therefore, exists only in the minds of individuals. From this view, he derived the epistemological thesis that all knowledge about society is based on knowledge about individuals.

Austrian methodological individualism achieved its most developed form in the writings of Friedrich von Hayek. His version is a synthesis of elements taken from Menger, Weber, and Mises. Like the

former, he suggests an atomistic method to study unintended consequences, but the terminology is different. For Hayek, it is a matter of using a "composite," or "synthetic," method to study "spontaneous orders." The influence from Weber and Mises is most clearly visible in his subjectivist view of social institutions and artifacts. Money, for instance, is constituted by the beliefs people entertain about the pieces of metal and paper used as money and not just by money itself. All social phenomena are subjective, and society as a whole is constituted by the beliefs of individuals.

Popperian Methodological Individualism

The philosopher Karl Popper is probably the single most influential advocate of methodological individualism: It puts forward the postulate that we must understand and explain all social phenomena in terms of the aims and beliefs of individuals, their action and interaction with one another, and the situation they are in. Popper takes exception to the psychologism of John Stuart Mill and supplements his own individualism with situational logic, which is a version of rational choice, and with institutionalism, which claims that the actions of individuals usually cannot be explained without reference to social institutions. So far, there is little to distinguish Popper's methodological individualism from that of the Austrians. They too make reference to social institutions in their models of social phenomena. There is one important difference, however. While the Austrians tended to treat institutions as phenomena to be explained, Popper sees them as part of the explanation. The difference becomes even greater in conjunction with the objectivism of Popper's philosophy. For Popper, institutions do not reside only in the minds of individuals but are also part of their environment, or situation.

Popper was particularly concerned to combat the Marxist and historicist kinds of holism. His methodological individualism gave rise to an intense debate in the 1950s, and it fell upon his pupils to clarify and defend it, but in different ways. J. W. N. Watkins, who wrote most about it, kept to the strictly individualistic principle, while Joseph Agassi and Ian C. Jarvie developed Popper's institutionalism and objectivism.

According to Watkins, the principle of methodological individualism states that social phenomena should be explained in terms of the principles governing the behavior of individuals together with a description of their situations. Unlike Popper, however, Watkins excluded institutions from the description of the situation. If they nevertheless pop up, they should be reduced to the attitudes of individuals toward things and other people. This is closer to the Austrians than to Popper, but Watkins shared the Popperian view that methodological individualism is about the explanation of social phenomena but not about the definition of collective concepts.

At first, Watkins did not distinguish methodology from ontology and epistemology, but eventually, he reached the conclusion that methodological individualism, strictly speaking, is a rule for social analysis, based on the ontological thesis that social phenomena are caused and constituted by individuals and the epistemological thesis that only individuals are observable. While we have direct access to facts about individuals, knowledge about social wholes is always derivative.

It may, finally, be noticed that Watkins does recognize the existence of holistic generalizations about large-scale social phenomena and comes to accept them as a legitimate part of social science. He still believes, however, that they are "in principle" reducible to psychological laws. These generalizations provide "half-way" explanations in terms of large-scale social phenomena, as distinguished from the "rock-bottom" explanations in terms of individuals and their interrelations, which are the ultimate goal of social science. With the recognition of half-way explanations, the status of methodological individualism changes from a rule to an aspiration.

With Agassi, we get an explicit statement of a new version of methodological individualism, which was only implicit in the writings of Popper. It is called "institutional individualism," and it derives from Popper's institutionalism and situational logic, rather than from his methodological individualism. The important point about institutional individualism is that it includes social institutions as objective elements in the situation of individuals. According to institutional individualism, social phenomena should be explained in terms of both individuals *and* social institutions.

The difference between the original versions of methodological individualism and institutional individualism was clarified by Jarvie, who suggested that methodological individualists typically put social wholes in the explanandum, while holists put

them in the explanans. Institutional individualists put both individuals and social institutions in the explanans.

Recent Developments

The debate about methodological individualism has waned somewhat since the 1960s but has never ended. In economics, confidence in methodological individualism seems less today than it used to be, and it is possible to see a shift from the original version to institutional individualism. In sociology, a new version, called "structural individualism," has emerged. It is akin to institutional individualism, but the main idea is that the explanation of social phenomena often requires reference to the positions occupied by individuals in social structures. In philosophy, methodological individualism is usually discussed as a form of reductionism and, recently, most often in light of the notion of supervenience.

Varieties of Methodological Individualism

Austrian methodological individualism was a principle about both the explanation of social phenomena and the definition of collective concepts. With Popper and his followers, it ends up as a rule only for the explanation of social phenomena. At times, methodological individualism has also been conceived as a doctrine suggesting the reducibility of social laws to (psychological) laws about individuals.

Of these three versions, the second has been the most common and also appears the most viable. Explanation is a more incontestable aim of science than is definition. It has also proved more difficult to define collective concepts than to explain social phenomena in terms of individuals. Considering the well-attested paucity of laws in social science, the reductionist thesis, too, seems problematic. If there are no laws, there is nothing to reduce.

In addition, the explanatory version of methodological individualism has the advantage of being closely associated with the plausible ontogenetic thesis that social phenomena are caused by the action of human beings.

Speaking literally, *methodological* individualism is a principle, or rule, telling social scientists how to proceed. Methodology is normative. Many advocates of methodological individualism, however, state it as a thesis about the nature and cause of social phenomena. Most common, however, is to use ontological individualism to support methodological individualism. Since ontological individualism is considered trivially true, methodological individualism follows as a matter of course.

Sometimes, methodological individualism is also stated as an epistemological thesis about knowledge. Since only individuals are directly observable in society, it is suggested that all knowledge about social phenomena can, at least in principle, be stated in terms of individuals.

On closer scrutiny, it is possible to identify a number of more or less radical versions of methodological individualism. They range from versions requiring a full explanation of social phenomena in terms of individuals to versions requiring only that individuals are part of the explanation, while assigning virtually all explanatory power to social institutions and social structure.

The main divide is between those versions that admit only individuals and their interaction in the explanans and versions that allow social institutions and social structure to be part of the explanation. In the former, social institutions and structures always appear as the results of interacting individuals. In the latter, the actions and interaction of individuals can be seen as the result of social institutions and structures. The former may be called *strong* and the latter *weak* versions of methodological individualism.

Lars Udehn

See also Austrian Economics; Causation in the Social Sciences; Emergence and Social Collectivism; Historicism; Holism, in the Social Sciences; Institutionalism and Institutional Theory; Microfoundationalism; Popper's Philosophy of Science; Reductionism in the Social Sciences; Situational Logic; Spontaneous Order; Supervenience

Further Readings

Agassi, J. (1960). Methodological individualism. *British Journal of Sociology, 11,* 244–270.

Brodbeck, M. (Ed.). (1968). *Readings in the philosophy of the social sciences.* New York, NY: Macmillan.

Hayek, F. A. von. (1942–1944). Scientism and the study of society. In F. A. von Hayek (Ed.), *The counter-revolution of science: Studies in the use and abuse of reason* (pp. 11–101). New York, NY: Free Press of Glencoe.

Jarvie, I. C. (1972). *Concepts and society.* London, England: Routledge & Kegan Paul.
Lukes, S. (2006). *Individualism.* Colchester, England: ECPR Press. (Original work published 1973)
Martin, M., & MacIntyre, L. C. (Eds.). (1994). *Readings in the philosophy of social science.* Cambridge: MIT Press.
Popper, K. (1945). *The open society and its enemies: Vol. 2. Hegel and Marx.* London, England: Routledge & Kegan Paul.
Udehn, L. (2001). *Methodological individualism: Background, history and meaning.* London, England: Routledge.
Watkins, J. W. N. (1957). Historical explanation in the social sciences. *British Journal for the Philosophy of Science, 8,* 104–117.

Induction and Confirmation

The Problem of Induction

As David Hume observed, first in his *Treatise on Human Nature* (1739) and then, more succinctly, in the *Enquiry Concerning Human Understanding* (1748), there is no contradiction in thinking that the future will be unlike the past in any way one cares to specify or in thinking that unobserved features of the world may behave very differently than those observed. But what philosophers call "the problem of induction" runs, as Hume himself made clear, much deeper than the mere failure of past history to logically entail any consequence about the future and the unobserved. Rather, past experience, exactly because it is experience of the past and the experienced, cannot tell us what features of our experience to date are so much as *relevant* to founding correct expectations about the future and the unobserved. No matter how well our plans and projections have worked in the past, no matter how much we may argue, in the modern vein, that without a huge amount of stability and continuity in the way things are and work, a species such as ours, able to reason and to use language, could not have evolved, all that concerns what we take ourselves to know about what has been the case up to now. We need, it seems, a principle that the future will resemble the past, a principle of *the uniformity of nature,* but we cannot hope to ground rational acceptance of the principle save by appeal to that very principle.

Hume's aim is not to impugn our everyday habits of thought, the more so, as we certainly do not have good grounds on which to believe, say, that the sun will *not* rise tomorrow, but rather to show us that they are no more than *habits* of thought, thoroughly well entrenched, psychologically inescapable habits, no doubt, but habits and nothing more. This "sceptical solution," as Hume called it, has been thought scandalous. Surely, it is said, some expectations about the future course of events, some projections from observed behavior to the behavior of things unobserved, are more reasonable than others (and in an everyday sense, some are, but reason in the narrower sense in which Hume uses it in setting up the problem does not, he says, make them so). From Immanuel Kant, roused, as he said, from his dogmatic slumbers by reading Hume, onward, much of philosophy—in part spurred by the success of (mostly natural) science—has aimed to explain how some beliefs about the future and the unobserved are justified, or at least better justified than others; but even if, with Peter Strawson, we say that our inductive practices are in part constitutive of rationality (and hence require no justification), we must recognize, as Strawson did, that they are not guaranteed success.

Varieties of Inductive Practices

Hume does not decry our inductive practices (how could he, since he thought that we cannot but employ them?), but what are they? The simplest is *enumerative induction,* reasoning from observed cases to unobserved ones (a process identified by Aristotle and criticized by Sextus Empiricus, ca. second to third century BCE, for reasons very like Hume's). The traditional example, which argued from all observed swans being white to all swans being white, came to grief when de Vlamingh sailed up the Swan in Western Australia and named it for its black swans. A similar pattern of reasoning involves statistical extrapolation from samples, projecting percentages in the sample to the population in general. Another involves curve fitting—why should we draw the smoothest line through plotted data points? Each requires care in use so that what is observed/sampled must be representative, to which end there are heuristics governing variety and randomness.

Nelson Goodman's "New Riddle of Induction" in 1946 showed matters to be more complex. Let *t* be some time in the future, and let *grue* mean "examined before *t* and green or not examined before *t* and blue." Let us suppose that all emeralds examined thus far have been green. Then all are, equally, grue. But the generalizations "All emeralds are green" and "All emeralds are grue," both obtained by the pattern of enumerative induction, disagree on emeralds not examined before *t* and so are inconsistent should there be any such. Clearly, we cannot reasonably accept the conclusions of both inferences. Goodman called those predicates that can be used in inductive inferences *projectible* and suggested that what makes "green" projectible, and "grue" not, is its use (and the use of cognate terms in other languages) in past actual projections (inductive inferences); "green" is *entrenched,* he says, whereas "grue" is not. Some have preferred to take the projectible predicates to be those that pick out *natural kinds,* objective kinds found in the natural world, not artificial or owing anything to our classificatory endeavors—the kinds that, to borrow from Plato, "carve nature at its joints." Theories in the natural sciences have as one of their aims, it is said, the determination of natural kinds—chemical elements provide the standard example. But if induction is used in the development or assessment of scientific theories, then circularity threatens.

Enumerative induction is often pictured as a means of arriving at a generalization and thus has a close connection with individual psychology and learning. As such, it has been modeled and has undergone refinement and prodigious development in artificial intelligence, wherein, in the subdisciplines of inductive learning, computational learning, algorithmic learning, and formal learning theory, we find notions such as the Vapnik-Chervonenkis dimension, Leslie Valiant's Probably Approximately Correct learning, and questions of computational tractability (computation in polynomial time). Two commonalities in this work are that (1) a means for encoding the data is given *ab initio* (circumventing "grue"/green issues) and (2) the hypothesis space is constrained; there is no modeling of *unbiased* learning, no computational *tabula rasa.*

For the American pragmatist Charles Sanders Peirce (1839–1914), *abduction* is the process by which we arrive at hypotheses. From available explanations, we choose that which best explains the phenomena in question. Brought to prominence in our times by Gilbert Harman in 1965 as a better account of what we do, even in cases of apparent enumerative induction, this has subsequently been turned into a very general account of the *confirmation* of theory by evidence: That theory is most likely to be true that scores highest on a balancing of various theoretical virtues, such as simplicity, explanatory power, consistency with the evidence, and coherence, both internal and with accepted theories. (There's a subtlety here: If "bestness" helps us arrive at the hypothesis/theory, then the initial plausibility may be a feature included in determining the best; if "bestness" is what is involved in confirming the hypothesis, its being the best explanation is what makes it the most plausible, given the available evidence.)

Confirmation

In the 19th century, John Herschel drew a distinction subsequently important for the philosophy of science, that between the *context of discovery* and the *context of justification* (in now standard terminology taken from Hans Reichenbach's *The Rise of Scientific Philosophy,* 1951). According to Herschel, how we come to a generalization or, indeed, any scientific hypothesis or theory (context of discovery) matters little for science and matters less the better and more plentiful the agreement with observation and experiment (context of justification). The idea that one should not just observe Nature at work but contrive tests goes back at least to the Middle Ages, to Roger Bacon (ca. 1214–1292). Three centuries later we find Francis Bacon (1561–1626) advocating the experimental method that was to be central to the foundation of the Royal Society in 1660. (Hume, it should be noted, conceived of his *Treatise* as "an attempt to introduce the experimental method of reasoning into moral subjects.")

Having drawn Herschel's distinction, inductive reasoning takes on a broader guise, that of the *confirmation* of generalizations, hypotheses, and theories by evidence. Enumerative induction can still play a role, but it now features in the support given to a generalization, say "All emeralds are green," by the observation of increasingly large numbers of green emeralds and no emeralds of any other color. The simplest and perhaps most obvious suggestion is *Nicod's criterion,* named for Jean Nicod

(1893–1924): that in the absence of an *A* that is not a *B*, the generalization "All *A*s are *B*s" is confirmed (supported) by the observation of *A*s that are also *B*s and by nothing else—confirmation by positive instances. By support or confirmation, we mean being positive evidence, evidence for, evidence in favor of a generalization or hypothesis. Doubt was cast on Nicod's criterion by Carl Hempel's *Raven Paradox*. Take the generalization "All ravens are black." By Nicod's criterion, this is confirmed/supported by the observation of black ravens. But logically speaking, "All ravens are black" says no more and no less than the generalization "All nonblack things are nonravens"—the two are logically equivalent (at least in classical logic, which is accepted without question in most methodological studies). By Nicod's criterion, this generalization is confirmed by the observation of anything that is neither a raven nor black. But, the thinking goes, logical equivalents should be confirmed/supported by the same evidence (for they are true and false in exactly the same circumstances), so the observation of a green emerald, say (i.e., a green emerald being a nonblack nonraven) confirms the generalization "All ravens are black." This conclusion is counterintuitive, hence the epithet "paradox." (Odder still, logically speaking, "All ravens are black" says the same as "All nonblack ravens are nonravens," which has *no* positive instances.)

Responses to Hempel's Paradox

There are two ways to respond to a paradox: (1) reject one of the premises on which it depends or (2) accept the conclusion and try to mitigate its counterintuitiveness. As there are two premises, this gives us three responses to Hempel's paradox, all three of which have had their advocates. Accepting the conclusion, one can say that a black raven provides better evidence/more support than a green emerald, and one can defend this by contrasting the small number of ravens with the large number of things that are not black. Rejecting one premise, one can deny that logical equivalents are supported by the same evidence; the difficulty here is explaining why how one expresses the generalization should matter to what counts as evidence in its favor, logical equivalents in effect making the same claim about the way things are. The third possibility is to deny Nicod's criterion (a course suggested by the lack of positive instances of "All nonblack ravens are nonravens" and by various less contrived examples in the literature).

Alternatives: Falsifiability and Probability

Can science do without induction? Karl Popper (*The Logic of Scientific Discovery,* 1959; first published in German in 1934) claimed so. The hallmark of the scientific is, for Popper, falsifiability. Science progresses by subjecting bold hypotheses to the test; we learn from our mistakes. We should use well-corroborated theories, for they have survived severe tests, but there is, says Popper, no inference to their likely future success. Just this absence of an inference strikes many as problematic for the technological application of scientific theories: It is not sufficient that we use a theory not (yet) known to be in error; we want to base applications on a theory most likely to be sufficiently accurate in the intended domain of application. (The so-called Duhem-Quine problem shows that falsificationism is less straightforward than this quick sketch may suggest. No sophisticated theory is tested in isolation; if nothing else, there is theory involved in the construction of experimental apparatus, the hypothesis that the equipment is performing as it should, and, perhaps, assumptions about the reliability of lab technicians or field workers. Thus, one is never obliged to lay a failed prediction at the door of the theory notionally under test.)

According to the *hypothetico-deductive* account of scientific theories, predictions of unknown occurrences and explanations of known ones are deduced from hypotheses and known "initial conditions." Hempel laid down various seemingly plausible criteria for confirmation of hypotheses by evidence in this setting, but as he himself noted, unless one finds a way to rule out logical gerrymandering, intuition is in error. For example, it may seem that a hypothesis is confirmed by a successful prediction and that when a hypothesis is confirmed so are its consequences, but these two claims lead to the consequence that any evidence confirms any statement whatever, for *e* is entailed by the conjunction *e* & *s*, as is *s*.

Well aware that enumerative inductions have gone awry in the past, one might think that in place of the bold "All *A*s are *B*s" we should posit the more cautious "Probably, all *A*s are *B*s" in the conclusion of an inference employing enumerative induction. That's one way to go: qualify the conclusion of

inductive inferences. Another is to qualify the inference itself. By this it is meant that, holding that some hypotheses are better supported by available evidence than others, one qualifies the inference from evidence to hypothesis. The standard framework for this latter practice is the orthodox theory of probability that we have from Blaise Pascal, Pierre Fermat, and Christiaan Huygens, which gained formal perfection in the 1930s in Andrey Kolmogorov's axiomatization (standard but not universal: L. Jonathan Cohen [1923–2006] championed a conception of probability he found in Francis Bacon's works). Assuming that $P(e)$, the prior probability of the evidence *e,* is nonzero, the conditional probability of the hypothesis h in light of that evidence, $P(h|e)$, is given by

$$P(h|e) = P(h \,\&\, e)/P(e).$$

$P(h|e)$ is the posterior probability of h, the probability of h in light of the evidence. The probabilities are often read epistemically, as the degrees of belief of a (idealized) rational agent. Under the rubric of *updating by Bayesian conditionalization,* $P(h|e)$ represents the degree of belief in h that the agent comes to have on learning that e (and learning nothing more). The stock-in-trade of Bayesian methodologists is to use this framework to explicate methodological insights (often made by those who have little time for the Bayesian/probabilistic approach). With that in view, we arrive at the *positive relevance* model of confirmation:

- e confirms h when $P(h|e) > P(h)$.
- e disconfirms h when $P(h|e) < P(h)$.
- e is evidentially irrelevant to/neutral with respect to h when $P(h|e) = P(h)$.

(An older terminology would have "infirms" for "disconfirms.")

This conception is not only found in much contemporary philosophy of science; in the theory of legal reasoning, we find the same conception of probative evidence, a conception enshrined in U.S. Federal Rule of Evidence 401.

Applying probability to inductive reasoning is not new. In the late 18th century, Pierre-Simon Laplace, taking the world to be 5,000 years old and modeling the rising of the sun as an unbroken sequence of heads obtained with a coin of unknown bias, arrived by means of his *Rule of Succession* at odds of 1,826,214 to 1 of its rising on the morrow.

Of course, we not only have the idea of evidence confirming and disconfirming, we can make comparisons: One piece of evidence may provide more support, may have, as they say in legal circles, more *probative value,* than another. Especially as we start from quantified values, the prior and posterior probabilities, this leads naturally to the project of measuring the amount of confirmation. If there is either confirmation or disconfirmation, there is a difference between the posterior and prior probabilities of the hypothesis. For any quantity, there are three common ways to represent differences in value: (1) the difference, as in, say, *three meters longer*; (2) the ratio, as in *twice as long*; and (3) the proportional difference, as in *30% longer.* In the case of confirmation, numerous measures have been proposed, of which the difference measure, the ratio measure (and its logarithm), and the "odds ratio" measure, $P(e|h)/P(e|\text{not-}h)$ (and its logarithm), have featured prominently in the literature. Debate as to their merits focuses on both their formal properties and how well they accord with intuitive judgments. (The odds ratio gets its name because it is equal to the ratio of posterior to prior odds; the logarithm of the odds ratio was used by Alan Turing [1912–1954] in his cryptographic work at Bletchley Park during World War II.)

While read epistemically, there is room for disagreement on the exact nature of the probabilities involved. Rudolf Carnap favored a logical interpretation of probability and developed a formal inductive logic in which the sentences of a formal, logical language, of the kind that served well in the development of formal and mathematical logic in the 20th century, are assigned values determined by the structure of the language. In his *Tractatus Logico-philosophicus,* Ludwig Wittgenstein had assigned a probability of ½ to elementary propositions and treated them as probabilistically independent, a formulation that allows no learning from experience (as with tosses of a fair coin). Initially, Carnap assigned probabilities on the model of ascribing equal probabilities to the different numbers of heads possible in a sequence of tosses.

Concluding Remarks

As Carnap and his coworkers labored on this project through the 1950s and 1960s, he drew back from fixing on one particular assignment of probabilities; this, Goodman's paradox, and, most important, the recognition that theories in sophisticated sciences are rarely couched in the terms applied in

the description of the phenomena that support them persuaded the majority of philosophers of science that there is no formal inductive logic, akin to formal deductive logic, underlying the confirmation of scientific theories. Philosophers of science reached this conclusion regarding (inductive) *confirmation* around the same time artificial intelligencers began to tackle computational implementation of inductive *learning and hypothesis generation.*

Peter Milne

See also Abduction and Inference to the Best Explanation; Bayesianism, Recent Uses of; Explanation, Theories of; Falsifiability; Hypothetico-Deductivism; Kinds: Natural Kinds Versus Human Kinds; Pessimistic Induction; Probability; Scientific Method

Further Readings

Carnap, R. (1962). *Logical foundations of probability* (2nd ed.). Chicago, IL: University of Chicago Press.

Dieks, D., Gonzalez, W. J., Hartmann, S., Uebel, T., & Weber, M. (Eds.). (2011). *Explanation, prediction, and confirmation.* New York, NY: Springer.

Gabbay, D. M., Hartmann, S., & Woods, J. (Eds.). (2011). *Handbook of the history of logic: Vol. 10. Inductive logic.* North-Holland, Netherlands: Elsevier.

Goodman, N. (1946). A query on confirmation. *Journal of Philosophy, 43,* 383–385.

Goodman, N. (1955). *Fact, fiction and forecast.* Cambridge, MA: Harvard University Press.

Harman, G. (1965). Inference to the best explanation. *Philosophical Review, 74,* 88–95.

Hempel, C. (1945a). Studies in the logic of confirmation: I. *Mind, 54*(213), 1–26.

Hempel, C. (1945b). Studies in the logic of confirmation: II. *Mind, 54*(214), 97–121.

Herschel, J. (1830). *Preliminary discourse on natural philosophy.* London, England: Longman.

Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian approach* (2nd ed.). Peru, IL: Open Court.

Hume, D. (1739). *Treatise on human nature.* London, England: Penguin Books.

Hume, D. (1748). *Enquiry concerning human understanding.* London, England: Collier.

Lipton, P. (1991). *Inference to the best explanation.* London, England: Routledge.

Maxwell, G., & Anderson, R., Jr. (Eds.). (1975). *Induction, probability, and confirmation.* Minneapolis: University of Minnesota Press.

Popper, K. (1959). *The logic of scientific discovery.* London, England: Routledge. (Original work published 1934 in German)

Strawson, P. (1952). *Introduction to logical theory.* London, England: Methuen.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus* (F. P. Ramsey & C. K. Ogden, Trans.). London, England: Kegan Paul.

INFERENTIALISM

This entry delineates the core elements of inferentialism. Recent inferentialist accounts of language and thought emphasize the importance of social practices of language users in terms of giving and asking for reasons, the essential normativity involved in inferences, the presence of commitments in all such activities, and the indispensability of others and of intersubjective ascriptions of entitlements in making inferences or ascribing them. Inferentialism underlines the sociality of language and thought rather than truth or representation, as in orthodox accounts, and recent versions of it prioritize language over thought.

Inferentialism is a theory concerning what constitutes the meaning of an expression or the conceptual content of a thought and authorized inferential relations between linguistic and between mental items, respectively. Drawing on themes from the contemporary analytic philosophers Wilfrid Sellars and Michael Dummett and from the earlier philosophical logician Gottlob Frege, the American philosopher Robert Brandom has recently given this theory its most explicit formulation and defense in his *Making It Explicit: Reasoning, Representing, and Discursive Commitment,* in *Articulating Reasons: An Introduction to Inferentialism* and in subsequent work. In this latter version, inferentialism has been drawn by Brandom close to some Hegelian themes.

Language, Thought, and Inference

Inferentialism explains meaning in terms of use, where *use* is specified in terms of the social practices of giving and asking for reasons. These practices institute the norms constitutive of language as a public possession. The meaning of a linguistic expression or the conceptual content of a thought is determined by its role in reasoning, the inferences that would authorize using that expression or having that thought, and the inferences that it, in turn, authorizes or commits one to. There is room for disagreement about the relative priority (epistemological

or metaphysical) of language and thought, but most current inferentialists prioritize language over thought.

Inference is fundamentally a form of speech activity involving sentences. Strictly, an inference is something we *do*—a transition from a (set of) uttered sentence(s) to another sentence or set thereof—though it would be a mistake to think that inferences must be actions. Utilizing Sellars's view that our concepts of intentional states are modeled on semantic concepts applied to utterances, a generalized notion of inference can then be extended to episodes of internal cognitive activity. Inferentialism thus appeals to pragmatists, who emphasize that thought and language must be understood as activities that enable us *to cope with the world*. Inferentialists take the functional notion of sentences providing rational support for other sentences to be the most basic notion in the semantic arena, more basic even than truth, which they tend to treat in a deflationary way. Note that an inferential role is not a purely causal role—it is normatively assessable as contributing to a good or a bad inference; it operates in "the logical space of reasons"—to use Sellars's famous formulation (i.e., in giving and accepting reasons, an operation fundamentally unlike causal action in physical space).

Inferentialism's Attack on Semantic Orthodoxy and Representationalism

Inferentialism takes seriously Frege's claim that sentences are the basic units of meaning. The meaning of subsentential expressions depends on the contribution those expressions make to inferences involving the sentences using them. A clear example of this approach is the definition of the truth-functional connectives ("and," "or," "not," "if . . . then," etc.) by means of introduction and elimination rules for each connective in "natural deduction" systems. The truth-functional "&" simply *is* the expression such that premises "A" and "B" license the conclusion "A & B" and the premise "A & B" licenses the conclusions "A" and "B."

This distinguishes inferentialism from current semantic orthodoxy, which takes a relation between terms and (sets of) objects; calls it "reference" or "satisfaction," to be the primitive factor in a meaning; and then constructs a definition of truth for the language Tarski-style (which arguably accounts for any remaining aspect of meaning). Inferentialists also reject the parallel "representationalist" paradigm in the philosophy of mind, which assumes the existence of a primitive *representation* relation between mental states and objects in the world. Inferentialists deny that there is any single particular, privileged relation of representation or correspondence between our concepts and objects in the world; we can speak and think about the world in virtue of our complex interactions with the world. These include *language-entry transitions,* such as observation reports in response to an irradiation of our sensory surfaces; *language-exit transitions,* such as announcing one's intention to shut the window followed by one's shutting it; and *intralinguistic transitions,* such as inferring from "smoke here" to "fire nearby." Strictly, only intralinguistic transitions are inferences, but inferentialists tend to see roles in entry and exit transitions as significant in constituting the meaning of numerous expressions.

Justificatory Relations and Commitment

Justificatory relations is a broad category; it certainly includes more than those inferences recognized by formal logic. Consider, for instance, the inference from "*x* is red" to "*x* is colored." This is obviously a good inference but is not valid in any formal logical system. Traditionally, logicians treat it as an enthymeme, an inference in which a premise is left out. The inferentialist denies that every good inference must be squeezed into the Procrustean bed of formal logic. Thus, inferentialists recognize a category of good *material* inferences beyond the class of formally valid inferences. Brandom argues at great length that formal logic must itself be seen as the discipline of *making explicit* (by allowing us to formulate in a distinctive vocabulary) a particularly significant subset of inferences without which language and conceptual thought would themselves not be possible. Brandom also accepts Sellars's suggestion that modalities express our endorsement of and *commitment* to certain inferences.

Thus, Sellars speaks of natural *laws* as "inference tickets"; the law of reflection, for instance, offers us a license to infer (ceteris paribus) from "The light impinged on the mirror at xB" to "The light was reflected from the mirror at xB." This accounts for the fact that laws support counterfactuals, whereas empirical generalizations do not. *Moral* laws are practical inference tickets.

Brandom, recently, has distinguished *weak inferentialism,* according to which "inferential articulation is a *necessary* element in the demarcation of the conceptual" (2000, p. 28); *strong inferentialism,* according to which "inferential articulation *broadly construed* is *sufficient* to account for conceptual content" (p. 28); and *hyperinferentialism*—"the view that inferential articulation *narrowly* construed is *sufficient* to account for conceptual content" (p. 28). The "inferential articulation" of a sentence concerns not just the inferences to and from that sentence made by a single subject but the inference licenses it gives *others* and those that *others* give it. Language is a distinctly *social* practice. The commitments, entitlements, and endorsements in terms of which language use is explicated are *intersubjective,* and language users are responsible to their linguistic community in what they say and what they thereby do. Brandom argues that holding a normative status is a matter of being held accountable to the inferences endorsed.

The Importance of Language

Inferentialists tend to draw a significant line between the cognitive states of those with and those without language. There may be some language-*like* internal processing in nonlinguistic beings, but there is clearly no actual inferential behavior strictly so-called in infralinguals. Furthermore, without language, the reflection that enables one to make explicit what one is doing in using language or in having thoughts seems unavailable.

Willem A. deVries

See also Given, Myth of the; Hegelianism and Contemporary Epistemology; Normativity; Pragmatism; Social Anti-Individualism and the Mental; Truth, Philosophical Theories of

Further Readings

Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment.* Cambridge, MA: Harvard University Press.

Brandom, R. (2000). *Articulating reasons: An introduction to inferentialism.* Cambridge, MA: Harvard University Press.

Dummett, M. (1973). *Frege's philosophy of language.* New York, NY: Harper & Row.

Frege, G. (1879). Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens [Concept script, a formal language, modeled upon that of arithmetic, of pure thought]. In J. vanHeijenoort (Ed.) & S. Bauer-Mengelberg (Trans.), *From Frege to Gödel: A source book in mathematical logic, 1879–1931* (pp. 1–82). Cambridge, MA: Harvard University Press. (Original work in German published in 1967 by Louis Nebert)

Sellars, W. S. (1953). Inference and meaning. *Mind, 62,* 313–338. (Reprinted in *Pure pragmatics and possible worlds: The early essays of Wilfrid Sellars,* by J. F. Sicha, Ed., 1980, Atascadero, CA: Ridgeview)

Sellars, W. S. (1957). Counterfactuals, dispositions, and the causal modalities. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science* (Vol. 2, pp. 225–308). Minneapolis: University of Minnesota Press.

Information Ethics

This entry presents the historical background and conceptual development of moral questions as well as of methodological and scientific issues comprising the ethics of information, viewing information and communication especially in its technological form.

Historical Background

Information ethics, understood as a philosophical discipline dealing with good and bad practices of human communication, has a long history going back, in the Western tradition, to the question of freedom of speech (*parrhesia*) in the Greek polis, dealing particularly with the Sophists' and Socrates's criticisms of the mores, principles, and concepts underlying communication in all its practical and theoretical dimensions. Plato's questioning of the written *logos* as well as Aristotle's careful analyses of the uses of language and his theory of rhetoric are forerunners of ethical thinking on the pervasive biases and power structures of communication. The invention of the printing press by Johannes Gutenberg around 1450 and the Reformation that profited from it challenged the legitimization of practices of interpretation and application of the Bible by the Roman Catholic Church, but implicitly it also challenged any kind of monopolization of knowledge and its production, storage, retrieval, distribution, and criticism by any human agent based on power. The questioning of

customary premodern communication practices culminated in the Enlightenment and its criticisms of political and religious censorship. Freedom of the press, the transformation of private libraries owned by the nobility and the Catholic Church into public property, and several encyclopedic projects aiming at critical and broad public access to knowledge were some practical corollaries of information ethics in the modern era.

Information ethics became a matter of concern particularly with the invention of the computer and the widespread use of the Internet since the last decade of the 20th century. Terms like *computer ethics* and *cyberethics* became popular. The term *information ethics* has been used particularly in relation to the information society by international agencies such as UNESCO since the end of the 1990s. This process culminated in the World Summit on the Information Society in 2003 and 2005.

Information Ethics as an Academic Discipline

The terms *information ethics* and *computer ethics* have been used since the early 1980s in the context of library and information science as well as computer science. In the mid 20th century, computer scientists such as Norbert Wiener (1894–1964) and Joseph Weizenbaum (1923–2008) raised ethical questions about the impact of computers on society, focusing on the responsibility of computer scientists. Computer ethics courses were introduced into academic curricula in the 1990s in schools of computer science in the United States. In the first decade of the new century, professional societies and networks such as the International Society for Ethics and Information Technology, the Centre for Computing and Social Responsibility (De Montfort University, United Kingdom), and the International Center for Information Ethics were set up. There is a large bibliography in the field, including specialized journals and handbooks as well as several international and regional conferences.

There are divergent views concerning the intention and extension of the concept of "information ethics"—that is, both its meaning and its reference. Some philosophers criticize the focus on human communication and plead for an extension to include all kinds of beings. This view turns eventually into informational metaphysics: questions of ontology and what to include in it. At the other end of the spectrum, there is the view of information ethics as dealing only with human communication as shaped particularly by information technology. In this case, the concept of information ethics excludes nondigital communication media still used today or prevalent in other epochs and cultures. The concept of information ethics might be extended to cover all kinds of information technology applications beyond the sphere of human communication. From a digital perspective that addresses all beings insofar as they can be digitized in order to be understood, information ethics is grounded in information ontology and concerns good practices of being-in-the-world in the digital age.

Key Topics in Information Ethics

The relation between information technology and ethical practices is twofold. On the one hand, it deals with the impact of information technology on good practices and their principles, while on the other hand, it has to do with the ethical reflection on information technology, which could be less reactive and more proactive with regard to the new societal challenges arising from new information and communication technologies. In both cases, information ethics has the task of discussing good practices and their principles with regard to either digital information technology or other media.

Information ethics deals with descriptive and critical issues in different cultures and epochs, giving rise to intercultural information ethics. This includes, but is not restricted to, the question of universal practices and principles. There might be agreement on universal declarations, but their interpretation and application might be different according to cultural traditions. An example of intercultural dialogue in information ethics is the discussion on the concept of privacy from a Western versus from a Buddhist perspective. Practical consensus might involve different reasons that are the object of analysis and criticism by information ethics. Deontological and utilitarian theories play a major role in information ethics, no less than do theories grounded in hermeneutics, analytical philosophy, critical theory, social theory, Marxism, postmodernism, and critical rationalism, to mention just a few.

Classical topics in information ethics are privacy, identity, trust, justice, intellectual property, cyberwar, the surveillance society, plagiarism, censorship, gender issues, and information overload.

Information ethics addresses the effects of the materiality of information technology on the environment caused by electronic waste and especially by its export to Third World countries. It also deals with the economic and political impact of information technology. Ethical analysis and critical evaluation of the global digital economy concern the relation between transparency, privacy, and secrecy, no less than issues of justice regarding access to and use of the Internet. Information technology in general, and in social networks in particular, plays a major role in the political development of societies. They might strengthen liberation movements and enable new forms of democratic participation, but they can be misused for oppression and exploitation as well. The vision of a people-centered, inclusive, and development-oriented information society, as proclaimed by the World Summit on the Information Society, outlines the object of ethical scrutiny and evaluation in order to develop reliable social conditions for trust, security, and transparency.

New technological developments such as ambient intelligence, human–machine symbiosis, neuroelectronics, affective computing, augmented reality, bioelectronics, the future of the Internet, cloud computing, and quantum computing are among the most relevant challenges for information ethics in the foreseeable future. The underlying philosophical debate concerns theoretical and practical prospects for human freedom and self-understanding in the digital age. Both issues cannot be divorced from the relation between humankind and the world, as well as between human and nonhuman life, taking into consideration the dangers and opportunities arising from their manipulation and transformation based on the uses and abuses of digital technology.

Rafael Capurro

See also Human–Machine Interaction; Information Society; Knowledge Society; Social Networks; Systems Theory; Technological Convergence; Technoscience and Society; Trust, Social

Further Readings

Bynum, T. (2008). Computer and information ethics. In *Stanford encyclopedia of philosophy.* Retrieved from http://plato.stanford.edu/entries/ethics-computer/

Capurro, R. (2006). Towards an ontological foundation of information ethics. *Ethics and Information Technology, 8*(4), 175–186. Retrieved from http://www.capurro.de/oxford.html

Capurro, R., Frühbauer, J., & Hausmanninger, T. (Eds.). (2007). *Localizing the Internet: Ethical aspects in intercultural perspective.* Munich, Germany: Fink.

Floridi, L. (2010). *The Cambridge handbook of information and computer ethics.* Cambridge, England: Cambridge University Press.

Froehlich, T. (2004). *A brief history of information ethics.* Retrieved from http://www.ub.edu/bid/13froe12.htm

Himma, K. E., & Tavani, H. T. (Eds.). (2008). *The handbook of information and computer ethics.* Hoboken, NJ: Wiley.

Hongladarom, S., & Ess, C. (Eds.). (2007). *Information technology ethics: Cultural perspectives.* Hershey, PA: Idea Group.

van den Hoven, J., & Weckert, J. (Eds.). (2008). *Information technology and moral philosophy.* Cambridge, England: Cambridge University Press.

Information Society

The term *information society* has become a common phrase, but it is used differently and sometimes vaguely by different people, especially those from different disciplinary or professional backgrounds. In its widest sense, it implies a society or culture in which information is a fundamental component of life. It could be argued that this is a condition of all human societies, even the least developed. In practice, however, this broad definition is usually narrowed to mean cultures in which organized knowledge (i.e., information) is consciously used to inform actions and decisions. Such information may be transmitted orally but is more often recorded in some retrievable form that might be graphic, but it is more often linguistic and numerical.

Since the 1990s—when the phrase became more common—it has usually been assumed that the information storage and retrieval systems are electronic. Consequently, the information society is often understood as meaning one in which networked computers and other electronic devices are essential instruments in all aspects of personal, professional, and public life. By extension, some then assume that the phrase also encompasses information technology

(i.e., computation and associated communications systems) rather than, or as well as, the social, economic, and political dimensions of information capture, storage, retrieval, and use. Given this lack of agreement about the phrase itself, it is better to perpetuate the vagueness and to emphasize that the focal point is information itself.

Information

In the social sciences (including for this purpose library and information sciences), information is normally understood to mean knowledge that has been organized. The organizational process is essentially intellectual, although it has a physical manifestation. Thus, to take a relatively simple example, the complex ontologies that underpin the classification systems (e.g., Dewey or Library of Congress) that have been used in libraries since the late 19th century are manifested as alphanumeric codes, which are then used to determine the order in which books are placed on the shelves. An even simpler example is this present encyclopedia, arranged in the order of the letters of the Latin alphabet, following certain conventions that are almost universally understood. At a deeper level, however, individual pieces of information (what computer scientists would call *bits*) are brought together in a way that adds value so that the whole is greater than the sum of its parts. Even a simple piece of information—that Columbus crossed the Atlantic Ocean in 1492—is actually derived from multiple individual pieces or bits; if the statement about Columbus is deconstructed, it will soon be recognized that understanding it requires a range of knowledge (e.g., about geography and chronology), which is shared between the originator and the user of the information. The construction of knowledge from information is perhaps most explicit in computing, where the bits represent the smallest reduction of the sum total of the information that is being recorded and processed.

The gap between information understood at this theoretical level and its societal application may seem to be very great. Indeed it is, and it is precisely for this reason that pragmatic and empirical approaches to the study of the information society are perhaps the most fruitful.

The Development of the Information Society

The formal history of information recording and retrieval begins with the invention of writing some 5,000 years ago and the development of media on which written matter could be stored and preserved. In Asia and Europe, at various times and in different ways, these innovations evolved into books and documents of the kind with which we are still familiar. Mechanical means of reproduction were developed independently in East Asia and Western Europe, but it was in the latter—by using pieces of metal that printed individual letters and symbols—that it was finally developed into the craft of printing in the 15th century. Within a relatively brief period of time (less than 100 years in some of the more sophisticated parts of Europe), printing had become an indispensable element in the organization of the state, of religion, of education, and of many aspects of private as well as public life. By the middle of the 19th century, when the printing processes themselves had been mechanized and industrialized, there was an ever-increasing flow of information in cultures that were approaching universal levels of literacy, benefiting from unprecedented prosperity, and developing complex technologies of communication. Some of these technologies were physical—the railroads, for example—but others were based on relatively recent scientific developments, such as the use of electricity for telegraphs and telephones. These wired networks provided a conceptual precedent for the Internet in the later 20th century.

By the time of the outbreak of World War I in Europe (1914), information was already being stored and retrieved through managed systems. While these seem crude 100 years later, the use of devices such as the "3 × 5" index card revolutionized the systematic storage and retrieval of information. In both business and government, information came to be seen as a vital commodity to be protected from competitors and to be made available only for the benefit of its owners. The extreme example was state secrecy, leading to the wartime development of espionage. But the same awareness of both the value and the use of information lay behind the development of mechanical devices such as adding machines and cash registers, which came to be widely used. Telephone networks—underdeveloped in Europe, although not in the United States, before 1914—were another beneficiary of war, and in the 1920s and 1930s, the broadcast media—radio followed by television—brought new sources of information into people's very homes. Social change followed quickly on the development of these systems and media. Businesses could be more tightly managed, and innovative systems of management

evolved, exemplified in the so-called Fordist production line in the rapidly growing automobile industry. Governments needed, and could both gather and store, vastly more information about citizens to support the social security systems that were developing in Europe and, by the late 1930s, in the United States. New forms of entertainment, some of them social (e.g., the movie theaters) and others domestic (e.g., broadcasting) began to change people's use of their leisure time and disposable income. And all the while, more people were better informed—and perhaps better able to exercise economic and political choice—than at any time in history.

Computers: The Information Machines

The computer was a wartime product. Building on work in theoretical physics and pure mathematics in the 1920s and 1930s, as well as experimentation in electromechanical engineering, a team of British scientists co-opted into the war effort produced the first true electronic computer. The commercial initiative in the development of computation crossed the Atlantic in the immediate postwar period, but not before the first machines had been installed in a handful of businesses. Early predictions that the world would need the power of two or three machines are easy to mock, but it was perhaps a decade before anyone outside the inner circles of the mathematicians, scientists, and engineers who were most closely involved really began to appreciate the potential of these new machines. The concept of a thinking machine—an electronic brain in the jargon of science fiction movies of the 1950s—came close to reality in machines that, even if they could not think, could certainly process data and information on a scale that the human brain could not. Although the earliest applications were almost entirely numerical, by the late 1960s, computers were in use for storing and sorting alphanumeric data, and input and output devices were being developed that made their use relatively easy. The period from 1970 to 1990 saw innovations that brought computing both economically and technically within the capacity of tens of millions of people; between them, the silicon chip and the mouse transformed an esoteric science into a tool in daily use.

The same two decades saw another advance: academic experiments in whether computers could be electrically linked—whether they could "talk" to each other in the popular language of the 1970s—were transformed into the infinitely complex and yet extremely simple-to-use Internet. Like computers themselves, the Internet had its origins in the military and intelligence communities. Like them, its civilian applications were transformative of whole societies. By the late 1990s, networked computing was commonplace in the academic world, was becoming normal in the U.S. government and was spreading into the wider world. The development of a simple interface, using the highly sophisticated concept of hypertext, was all that was needed to propel the final stage of takeoff. The World Wide Web, emerging from the civilian scientific community among theoretical physicists in Europe, became the platform for the near-universalization of the Internet and in the eyes and minds of billions of ordinary men and women (and above all children and young people), became the Internet itself. By the turn of the millennium, the Internet was on the verge of becoming mobile. Cellular telephones presaged the independence of devices from wired networks, and as devices became both cheaper and more powerful, the mobile telephone became ubiquitous; indeed, in some parts of the world, where telephone services had always been unreliable, mobile telephony is probably the most important single development in the history of communication. By the end of the first decade of the 21st century, the cell phone had become a mobile computer providing access to the Internet, and hence to e-mail, as well as to voiced telephony and text messaging systems, the latter the preferred medium of the coming generation. Superimposed on all of this—and to some extent driving it—were the search engines that actually provided more or less systematic access to the ever-increasing information content of the Internet.

A New Society?

It is these two and a half decades of change that are argued by some to have made the information society a reality. But others take the view that we should not fall into the trap of technological determinism. It is certainly the case that cheap and mobile access to networks and all the content they can carry, as well as the interpersonal communication systems that they facilitate, has changed how people live and work. There is a far greater consciousness of information, and perhaps a greater understanding of it. On the other hand, the roots of this information society lie deep in Western culture, with its

tradition of literacy and text, which goes back more than 2,000 years, and a 1,000-year-old system of education, beginning with the acquisition of literacy. Governments have been seeking, collecting, and protecting information for centuries in surveys, mapmaking, and population censuses. Business owners have kept financial records for at least as long, even if it was only relatively recently (say in the past 150 years) that such record keeping has been standardized. Information and computing technologies have made all of this easier and indeed have made a practical reality of systems that could previously only be conceptualized. Some social scientists have argued that all that has happened is that there are new and infinitely faster and generally more effective tools for doing old things. Others prefer the view that the change is as profound as it has been rapid. Certainly it is radical. By the middle of the 21st century, the world will be led by men and women who have never known any other society than the one we have now, people to whom the mobile information and communication devices are as normal as automobiles (and perhaps by then more socially acceptable) and people to whom the transmitted word is a manifestation of an electronic file, not the physical product of a printing press.

The social and economic benefits of the information society are clear enough, but even many of those who accept that concede that there is another side to the coin. Ease of electronic communication, whether by voice, text, or e-mail, can be cogently argued to have changed human interaction and diminished the importance of face-to-face meetings and conversations. Colleagues who once visited each other's offices or used the telephone now work by text messages or e-mail. Social networking systems allow members to have scores or hundreds of "friends" whom they have never met, and about whom they actually know nothing or may "know" only the deliberate falsehoods that the "friend" has posted. This may be thought to be socially or psychologically disturbing. Even more disturbing perhaps, is that the information society can be readily transmuted into a surveillance society. Governments, even in democracies, have had some powers to tap telephones or open mail, but such powers were carefully regulated by legislation and usually required the specific permission of a court of law. Electronic communications are easier to tap into—which journalists know well and exploit to their advantage—and it can be done without either sender or recipient being aware of the fact. The use of closed-circuit television systems in streets, in buildings, and on public transportation systems is redefining the very concept of public space and indeed of privacy itself. Legislation has, generally, lagged behind the rapid technical change, and politicians have shown a tendency either to be ignorant of the issues or to consciously ignore them. The regulation of the information society—striking the democratic balance between the rights and obligations of the citizen and those of the state—is an issue that urgently needs to be addressed.

John Feather

See also Encyclopedia; Experimenting Society, The; Information Ethics; Knowledge Society; Philosophy of Expertise; Technological Convergence

Further Readings

Black, A., Muddiman, D., & Plant, H. (2007). *The early information society: Information management in Britain before the computer.* Aldershot, England: Ashgate.

Castells, M. (1996–1998). *The information age: Economy, society and culture* (3 vols.). Oxford, England: Blackwell.

Feather, J. (2008). *The information society: A study in continuity and change* (5th ed.). London, England: Facet.

Gleick, J. (2011). *The information: A history, a theory, a flood.* London, England: Fourth Estate.

Naughton, J. (1999). *A brief history of the future: The origins of the Internet.* London, England: Weidenfeld & Nicolson.

Webster, F. (2007). *Theories of the information society.* Hoboken, NJ: Taylor & Francis.

Institutional Economics

Institutional economics was a term originally used to describe the works of a prominent group of American economists including Thorstein Veblen, Wesley C. Mitchell, John R. Commons, and many others. Although its influence declined, even after World War II the original institutionalism retained some prominence. Simon Kuznets and Gunnar Myrdal declared an affinity with the original institutionalism and won Nobel Prizes in 1971 and 1974,

respectively. In 1975, Oliver Williamson coined the term *new institutional economics* to apply to work by him and others, including Ronald Coase, Mancur Olson, and Douglass North. These authors were reviving interest in the role of institutions in the economy, and much of their work used the concept of *transaction costs*. Since then, Ronald Coase, Douglass North, Elinor Ostrom, and Oliver Williamson have all been awarded Nobel Prizes for their work in institutional economics.

This entry gives an overview of ideas in the original and new traditions of institutional economics.

A Prominent Feature of the Original Institutional Economics

The original institutionalism was highly diverse, but it did exhibit some common themes. Among these was the emphasis placed on the role of institutions in both understanding the economy and developing policies for social reform. Another common theme pervades institutionalism, from the writings of Veblen in the 1890s to those of John Kenneth Galbraith, who died in 2006—the notion that the individual is not given but can be reconstituted by institutions. By contrast, much of mainstream economics during the 20th century, including much (but not all) of the new institutional economics, takes individual preferences as given.

The assumption of malleable preferences is often criticized as leading to some kind of structural or cultural determinism or to methodological collectivism. The individual, it is said, is made a puppet of institutional or cultural circumstances. But in the writings of Veblen and Commons, there are both upward and downward effects; individuals create and change institutions, just as institutions mold and constrain individuals. The old institutionalism is not necessarily confined to the "top-down" cultural and institutional determinism or methodological collectivism with which it is sometimes associated.

A merit of the idea that institutions shape individual dispositions is that it admits an enhanced concept of *power* into economic analysis. Power is exercised not only by forceful coercion. Power is often exercised more cleverly—and often without overt conflict. Supreme power is exercised by subtly influencing the thoughts and desires of others. These considerations are almost entirely absent from mainstream economics.

Preference malleability is also important in regard to *learning*, which typically takes place through and within social structures. Mainstream economics has difficulty accommodating a full notion of learning because the very idea of "rational learning" is problematic. It also assumes that our preference function is at birth fully primed to evaluate options about which we are not yet aware and that may not have emerged. But instead of the mere informational input of "facts" to given individuals, learning in practice is a developmental and reconstitutive process. Learning involves adaptation to changing circumstances, and such adaptations mean the reconstitution of the individuals involved. Institutions and cultures play a vital role in establishing the concepts and norms of the learning process.

Because conceptions of social power and learning are placed at the center of economic analysis, the original institutionalism is arguably more suited to address questions of structural change and long-term economic development, including the problems of less-developed economies and the transformation processes in the former communist countries. On the other hand, the analysis becomes much more complicated and less open to formal modeling. In normative terms, the individual is no longer taken as the best judge of his or her welfare. A different type of welfare analysis is required, but such alternatives are so far underdeveloped in institutional economics.

The Evolving New Institutionalism

Since its development by Coase and Williamson, the concept of *transaction costs*—defined as the costs of formulating, monitoring, and enforcing contracts—has generated a vast theoretical and empirical literature and helped us understand the nature and operation of firms and markets. But comparisons of different institutional structures in most transaction cost analyses do not consider the possibility that individual preferences are molded by different institutional circumstances.

The "new institutional economics" originally set out to explain the existence of political, legal, or social institutions by reference to a model of given, individual behavior, tracing out its consequences in terms of human interactions. The explanatory movement is from individuals to institutions, taking individuals as primary and given.

But the reception of information by individuals requires cognitive norms and frames to process and make sense of that information. Furthermore, our interaction with others requires the use of the institution of language. Original institutionalists argued that the transmission of information from institution to individual is impossible without a coextensive process of *enculturation*, in which the individual learns the meaning and value of the sense-data that are communicated.

Institutions constrain, enable, and influence individuals. Accordingly, if there are institutional influences on individuals and their goals, then these are worthy of explanation. At the very minimum, analysis of the development of institutions depends upon interpersonal communication of information. And the communication of information itself requires shared conventions, rules, routines, and norms.

This does not mean that new institutionalist research is without value, but it suggests that the starting point of explanations cannot be institution-free: The main project has to be reformulated as just a part of a wider theoretical analysis. The reformulated project would stress the evolution of institutions, in part from other institutions rather than from a hypothetical, institution-free "state of nature." It also means that the existence of the firm may not be explained by transaction costs alone; other causal factors may also be relevant.

Some "new" institutionalists have taken these points on board. Douglass North has insisted on the general importance of understanding the context and processes of cognition. He has cautioned on the limits of the rational choice framework, emphasizing the way institutionalized ideologies influence individual cognitions and goals. This places North's recent writing close to the original institutionalist tradition.

Other contemporary institutionalists have abandoned the idea of starting from given individuals alone. For example, Jack Knight criticizes much of the new institutionalist literature for neglecting the importance of distributional and power considerations in the emergence and development of institutions. Elinor Ostrom emphasizes the role of culture and norms in establishing and molding both perceptions and interactions. Masahiko Aoki identifies the problem of infinite explanatory regress in much of the former literature and develops a novel approach. Using game theory, he explores the evolution of one set of institutions from another. These developments reveal some convergence between elements of the "new" institutional economics and ideas that can trace their origin to the original institutionalism.

The Future Development of Institutional Economics

Once we move in the direction of a more open-ended evolutionary approach, another question is raised. If in principle every component in the system can evolve, then so too can individual preferences. Institutions involve rules, constraints, practices, and ideas that can—through psychological and social mechanisms that need to be revealed—sometimes mold individual purposes and preferences in some way.

Such intuitions can be found in the original institutionalism. But what is lacking in some of this literature is a clear exposition of the causal processes involved. One of the most fertile explanations of the processes is found in the writings of Veblen, where he describes the influence of selective competition between institutions on individual *habits*. Veblen took the concept of habit from American pragmatist authors including Charles Sanders Peirce and William James. Habits are acquired psychological dispositions to respond in particular ways to specific triggers. Insofar as individuals are constrained or motivated to follow particular institutional norms or rules, they tend to strengthen habits that are consistent with this behavior. Individuals may then rationalize these outcomes in terms of preference or choice. That is, conscious preferences are the outcome of habits, rather than the other way round. Our habits help make up our preferences and dispositions. When new habits are acquired or existing habits change, then our preferences alter. The framing, shifting, and constraining capacities of social institutions give rise to new perceptions and dispositions within individuals. Once habits become established, they become a potential basis for new intentions or beliefs. As a result, shared habits are the constitutive material of institutions, providing them with enhanced durability, power, and normative authority.

Alongside Veblen, the contribution of Commons to the interface of economics and law is still relevant. Veblen was theoretically weak on the nature and analysis of market institutions, and here we must learn from others. A number of institutionalists have taken up Keynesian ideas, but so far only limited progress has been made in developing robust

theoretical links between the theoretical analysis of institutions and macroeconomic theory. Finally, all institutional approaches so far lack a developed alternative grounding for economic policy to rival the patently limited neoclassical welfare approach, where the individual is assumed to be always the best judge of his or her interests.

Much work in the New Institutionalism is oriented toward practical policy design. Those influenced by the original institutionalism are also interested in policy, but they are also disposed to be more reflective about the philosophical foundations of their approach. These concerns have become prominent in recent discussions about the nature of institutions. Although there is widespread agreement that institutions are systems of rules, there is ongoing discussion concerning the nuances of this definition and how broadly the term *institution* may apply. Such discussions involve social ontology as well as other philosophical issues, as exhibited in the work of John Searle, among others.

Conclusion

There is now much overlap, and the possibility of fruitful dialogue, between original and New Institutionalism. There has also been an "institutional turn"—recognizing the importance of institutions—in other social sciences, including politics, sociology, history, and geography. What emerges as institutional economics in the next few decades may turn out to be very different from what was prominent in the 1980s and 1990s, and it may trace its genealogy from the original as well as the New Institutionalism.

Geoffrey M. Hodgson

See also Heterodox Economics; Individualism, Methodological; Institutionalism and Institutional Theory; Institutions as Moral Persons; Philosophy of Economics, History of; Pragmatism and the Social Sciences

Further Readings

Aoki, M. (2001). *Toward a comparative institutional analysis*. Cambridge: MIT Press.

Brousseau, É., & Glachant, J.-M. (Eds.). (2008). *New institutional economics: A guidebook*. Cambridge, England: Cambridge University Press.

Coase, R. H. (1988). *The firm, the market, and the law*. Chicago, IL: University of Chicago Press.

Hodgson, G. M. (2004). *The evolution of institutional economics: Agency, structure and Darwinism in American institutionalism*. London, England: Routledge.

Hodgson, G. M. (2006). What are institutions? *Journal of Economic Issues, 40*(1), 1–25.

Knight, J. (1992). *Institutions and social conflict*. Cambridge, England: Cambridge University Press.

Ménard, C., & Shirley, M. M. (Eds.). (2005). *Handbook of new institutional economics*. Berlin, Germany: Springer.

North, D. C. (2005). *Understanding the process of economic change*. Princeton, NJ: Princeton University Press.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge, England: Cambridge University Press.

Rutherford, M. H. (2011). *The institutionalist movement in American economics, 1918–1947: Science and social control*. Cambridge, England: Cambridge University Press.

Schmid, A. A. (2004). *Conflict and cooperation: Institutional and behavioral economics*. Oxford, England: Blackwell.

Schotter, A. R. (1981). *The economic theory of social institutions*. Cambridge, England: Cambridge University Press.

Williamson, O. E. (1985). *The economic institutions of capitalism: Firms, markets, relational contracting*. London, England: Free Press.

Institutionalism and Institutional Theory

This entry gives an extended overview of an area that, in its recent rich revival, has brought back the study of institutional structures and organizations to the center of social sciences and, in particular, of political science. Institutionalism poses significant questions of epistemological and, more generally, philosophical import for the study of the social world, not least of which is the core antithesis between holism and individualism.

Origins

The study of institutions was the foundation of political science. Theorists going back to the greats such as Aristotle, Baron de Montesquieu, Thomas Hobbes, and John Locke (among others) sought to understand how institutions function and how to

design better institutions. More recently, much of comparative politics was based on the study of public-sector institutions and constitutions, with the fundamental assumption that formal structures would determine what governments actually did as they attempted to govern.

As political science moved away from its organizational and institutional foundations toward assumptions about the centrality of individual behavior, then much of that central role for institutions was lost. First with "behavioralism" and then with rational choice theory, methodological individualism came to dominate political science, and institutions were at best merely structures within which individuals interacted. Beginning in the 1980s, however, there was a significant revival of interest in institutions and institutional theory, usually discussed as the "New Institutionalism."

The New Institutionalism

The New Institutionalism in political science differed from the older versions of this basic approach in several important ways. The most important is that the New Institutionalism is explicitly theoretical, attempting to develop more or less comprehensive explanations for political behavior on the basis of institutions. Furthermore, the New Institutionalism has been informed by the understandings about individual political behavior that have emerged from other parts of political science. Therefore, one of the central questions in contemporary institutional theory is how institutions shape individuals while at the same time individuals shape institutions. Finally, the New Institutionalism has extended its analytic reach from just issues of comparative politics to influence a wide range of topics such as international relations, political parties, and interest groups.

One Institutionalism or Many?

Although the revival of institutional theory has been successful in returning the concern for political structures and organizations to the center of political science, it was perhaps too successful in that it has spawned a wide range of alternative versions of institutional theory. Each of these alternative approaches provides important perspectives on the behavior of institutions and on politics more broadly, but each also has important limitations. Although a brief discussion of each approach cannot do them justice, it is important to at least identify the more important characteristics of each.

Normative Institutionalism

The first major version of institutionalism to emerge was normative institutionalism. It is here referred to in this manner because it assumes that the behavior of individuals within an institution is shaped by a "logic of appropriateness," meaning the norms, values, symbols, and myths within that institution. This perspective is in contrast to a "logic of consequentiality" in which individuals choose their actions based on their predictions and evaluations of the outcomes. In the normative perspective on institutions, individual preferences are endogenous to the institution, being created as individuals join the institution, and the process of socialization of new members is crucial for maintaining appropriate behavior within the institution.

The normative institutionalism is closely allied with strands of institutionalism in sociology, as well as closely connected to organization theory. Its emphasis on the norms and values defining institutions is similar to the approach of sociologists such as Philip Selznick, who have also emphasized the importance of values in institutions. This work also is in the tradition of the "bounded rationality" of Herbert Simon, who has argued that individuals cannot be comprehensively rational but rather operate within a set of boundaries on that rationality, notably those imposed by organizational memberships. The "logic of appropriateness" establishes a clear set of bounds on the rationality of individuals and provides the individual with a set of boundaries for behavior.

Rational Choice Institutionalism

Rational choice institutionalism is the second widely used approach to institutionalism, but it is almost diametrically opposed to the logic of normative institutionalism. Rational choice institutionalism assumes that individuals within institutions still attempt to maximize their personal utility, but they do so within the context of the incentives and rules that can be used to characterize institutions. In this perspective, institutions are defined by individual reactions to those rules and incentives so that the institution constitutes in essence an ecology of action for the individual participants.

In short, in rational choice institutionalism, individuals make their decisions according to a logic of consequentiality.

Even more than in most other approaches to institutions, in the rational choice perspective, institutional designs are undertaken explicitly to overcome certain fundamental social problems. One is the need to overcome collective action problems and to create equilibrium under conditions that otherwise would produce instability. Somewhat similarly, this approach to institutions is concerned with the capacity to overcome collective action problems, such as the tragedy of the commons, and problems of common pool resources. Finally, rational choice institutionalism focuses on the designs of institutions that facilitate or hinder action in government in general.

Historical Institutionalism

The simplest prediction for any social or political setting is the persistence of the status quo. In the historical institutionalism, this basic observation has become central to a theoretical approach to political science. The assumption behind the historical institutionalism is that the decisions about policies and organization made at the "formative moment" of an institution are likely to persist until they are transformed by some force sufficient to overcome the inertia in the system. This "path dependence" would persist until some "punctuated equilibrium," meaning major departures from the established patterns based on some fundamental transformation of the environment or of the ideas governing the policy domain.

Although historical institutionalism has been widely used within political science, it has clear roots in economics. Economists have been interested in why suboptimal solutions to problems persist, even in the presence of better options. The fundamental answer offered by economists has been the transaction costs of moving to superior solutions. Political scientists, on the other hand, have tended to focus on the reinforcements offered for continuing the existing solutions, especially the benefits that may be offered to political and administrative elites responsible for the programs. In this perspective, significant change in an institution will not occur without conflict over the nature and the mission of the institution.

Discursive Institutionalism

Most versions of institutionalism are concerned directly with structures and their interaction with the members of those institutions. Scholars have, however, developed an approach to institutions that depends more on the ideas of the members of those institutions. The logic of this approach is that members of an institution have alternative discourses about the nature of the institution and about its programs. These discourses may be coordinative, linking the members of the institution, or they may be communicative, linking the institution with its environment. In either case, however, the argument is that the best means of understanding the institution is to understand its discourses.

This discursive version of institutionalism is in many ways closely related to the normative institutionalism discussed above. Both of these approaches are based on ideas, with the assumption that those ideas rather than formal hierarchical structures are central in controlling the behavior of the members of the organization. One principal difference between the two approaches is that the normative version tends to assume that the institution strives toward one common set of ideas and that those ideas will be relatively stable. The discursive model, on the other hand, assumes more competition among competing discourses, so that any set of ideas will generally be in competition with alternative discourses.

Central Theoretical Questions

As well as the numerous differences existing among the several approaches to institutionalism, there are some important theoretical questions that cut across the approaches. The various approaches all provide somewhat different answers to these questions, but all must address these points if they are to function as effective theoretical approaches to institutions and to political life more generally.

Institutional Change

Perhaps the most difficult question for any institutional theory is how it integrates change into its framework for analysis. The basic logic of institutions is to create predictability and stability for individual behavior and within the structures themselves. In general, therefore, institutional theories tend to be better at explaining and predicting stability than they are at explaining change. Indeed, for historical

institutionalism, the principal focus of the approach is demonstrating the difficulties of change actually occurring in institutions, except in the most extreme circumstances. Although institutional theory does emphasize stability, we do know that institutions change, and we need to bring change into the theoretical apparatus.

Even the historical institutionalism that places stability at the center of its analysis of institutions must somehow find the means for understanding and coping with change. As noted earlier, the original presentations of the approach depended upon major changes—punctuations—as descriptions of changes, although there was little means of explaining those major shifts. More recent theorizing has retained the underlying logic of historical institutionalism while also admitting a number of more gradual mechanisms for change, such as layering or displacement.

At the other end of the spectrum of mutability, rational choice institutionalism implies the greatest facility for change. Since the preferences of individuals are unchanged by their involvement with the institution, all institutional change requires is to alter the rules and/or incentives for change to occur. If the incentives are designed properly then those who would alter institutional behavior can produce the types of behaviors that are desired. That said, scholars working in this tradition have also begun to consider less rationalistic mechanisms such as learning as sources of change.

The models of institutions based on values pose particular problems for change. While the internal control of individuals within an institution is facilitated by their normative commitment to the institution, producing change requires undoing efforts at having those members accept the "logic of appropriateness." If the leadership of the institution seeks to alter those beliefs, then resocialization and recommitment become necessary. This is not only difficult and time-consuming but may over time breed cynicism among members when they are given seemingly mixed signals over what constitutes "good" behavior within the institution.

Institutionalization and Deinstitutionalization

As well as being concerned about the way in which institutions change, institutional theorists must be concerned about how institutions form and how they disintegrate. The process of creating institutions, or institutionalization, is crucial for the success of an institution, but the conceptualization of this process is closely allied with the various approaches to institutionalism: Normative institutionalists, for example, tend to see institutionalization as infusing a structure with values, while historical institutionalists tend to assume that institutionalization occurs virtually automatically during the "formative moment" of the program.

Most institutional theory tends to consider institutions in a rather dichotomous manner—they either exist or they do not. In reality, however, institutions are always becoming more or less institutionalized. For example, from a normative perspective, if the dominant values of an institution are challenged, or are not being sufficiently followed by the members of the institution, then the institution will be to some extent deinstitutionalized. If the processes of deinstitutionalization go too far, then the institution will fail or perhaps be replaced by a new one in some process of "punctuated equilibrium."

Individuals and Institutions

Finally, the interactions of institutions and individuals are important for defining the manner in which institutions perform their tasks. A fundamental paradox dwells within this relationship in institutional theory. Institutions are human creations, but once created, they are meant (at least in some models of institutions) to shape human behavior and to be virtually immutable. Although seemingly paradoxical, this points to a crucial characteristic of institutional theory: Individuals shape institutions, but institutions also shape individuals.

Again, the manner in which this fundamental characteristic of institutional theory is actually manifested depends upon the nature of the particular approach to institutions. The normative approach, for example, emphasizes the way in which institutions shape the preferences and values of its members. Although less explicit, the historical institutionalism also assumes that being within an institution tends to shape the perspectives of individuals so that they will maintain the policy approach of the institution.

Individuals not only shape institutions at their inception, but they may have continuing influence over those institutions. Individuals are the raw material of public institutions—both as employees and

as the clients of their programs. If either of these two groups change in any fundamental way, then too the institution will have to adapt. For example, the changes in the recruits coming into the U.S. military during the Vietnam War required fundamental changes in military management, and those changes persisted in the institutions.

Conclusion

The preceding discussion has considered institutionalism in terms of the various alternative versions, all of which have something to say about institutions. Each of these approaches has something interesting to say, but the question that remains is whether there is enough in common among them to argue that there is an institutional approach, rather than just a collection of rather disparate approaches. While the alternative approaches are useful in their own right, and can illuminate many aspects of political behavior, if institutionalism is to be an alternative paradigm for political science, then there should be some greater integration of the ideas within these various versions.

At a minimum, all these approaches are concerned with creating regularities of behavior, especially in situations in which one would expect more diverse patterns. The approaches may generate that regularity in rather different ways, but they all still create greater regularities and greater predictability than would be found in social systems. Furthermore, the degree of regularity expected in the several approaches may differ significantly—with the discursive model requiring rather little—while some regularity is central to all the approaches.

Institutional theory is an important contribution to our collective understanding of social and political questions, but it is certainly not without its shortcomings. The most obvious problem is the lack of uniformity and common definitions of even basic points such as what constitutes an institution. But emphasizing that politics occurs largely within institutions, and that indeed a good deal of politics is the interaction of institutions, these approaches taken together have helped alter the perspective of the social sciences. The extreme individualism of some approaches to political life has been challenged, and the need to understand institutions has been returned to a central place in the social sciences.

B. Guy Peters

See also Holism, in the Social Sciences; Individualism, Methodological; Institutional Economics; Institutions as Moral Persons; Path Dependence; Performative Theory of Institutions; Rational Choice and Political Science; Social Choice Theory; Social Institutions

Further Readings

Katznelson, I., & Weingast, B. (2005). *Preferences and situations: Points of intersection between historical and rational choice institutionalism.* New York, NY: Russell Sage Foundation.

March, J. G., & Olsen, J. P. (1984). The new institutionalism: Organizational factors in political life. *American Political Science Review, 78,* 738–749.

March, J. G., & Olsen, J. P. (1989). *Rediscovering institutions: Organizational factors in political life.* New York, NY: Free Press.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action.* Cambridge, England: Cambridge University Press.

Peters, B. G. (2012). *Institutional theory in political science: The new institutionalism.* London, England: Continuum.

Peters, B. G., Pierre, J., & King, D. S. (2005). The politics of path dependency: Political conflict in historical institutionalism. *Journal of Politics, 67,* 1275–1300.

Riker, W. H. (1980). Implications from the disequilibrium of majority rule for the study of institutions. *American Political Science Review, 74,* 432–446.

Schmidt, V. A. (2010). Taking ideas and discourse seriously: Explaining change through discursive institutionalism as the fourth new institutionalism. *European Political Science Review, 2,* 1–25.

Selznick, P. (1996). Institutionalism "old" and "new." *Administrative Science Quarterly, 41,* 270–277.

Streeck, W., & Thelen, K. (Eds.). (2005). *Beyond continuity: Explorations in the dynamics of advanced political economies.* Oxford, England: Oxford University Press.

Tsebelis, G. (2002). *Veto players: How political institutions work.* Princeton, NJ: Princeton University Press.

Institutions as Moral Persons

This entry presents an important issue in social ontology: whether formal institutions, such as organizations or corporations, can be considered as bona fide moral persons. The entry discusses the principal theories' pros and cons.

The Issue

It is difficult to ignore the fact that institutions—formal organizations that evidence corporate managerial structures and decision making (unlike constructs of social order such as the family or marriage)—play dominant roles in the social, political, and economic spheres and that they are often targets of moral evaluation in ordinary discourse.

Nonetheless, since at least as early as the 17th century, it was fashionable in the philosophical literature to regard the actions and interests of institutions, for moral purposes, as reducible to the actions and interests of those humans who function within them. Only humans, it was held, can be moral persons and morally appraisable. "Moral persons," on that traditional account, are solely humans who are capable of intentional action, of conforming to rules, of understanding the moral significance of what they do and how they are perceived by others, and, in some measure, of being receptive and reactive to moral reasons to do one thing rather than another. That is, moral persons must be functionally intentional, responsive to moral reasons, and capable of affectivity. Opponents of treating institutions as moral persons maintain that institutions qua institutions cannot meet all or some of those criteria and so must not be admitted into the moral community.

Denying That Institutions Can Be Moral Persons

Two theories, one metaphysical (ontological) and the other meta-ethical, typically are intertwined in the position held by opponents of treating institutions as moral persons. The metaphysical one is usually referred to as *methodological individualism*. The other might be called *moral individualism*.

Methodological individualism holds that the behavior of an institution is always *reducible* to the behavior of the individuals who make up the institution (without remainder). Methodological individualists maintain that apparent institutional actions always can be reduced to a set of facts about individuals, which can then be arranged to provide a complete description of the behavior of institutions. For them, it might be said, the names of institutions are collective, not singular, nouns, and an institution is little more than a contractual nexus of humans. Many sociologists have rejected that theory because it does not adequately explain a significant number of social facts concerning organizational/institutional behavior.

Moral individualists, who may or may not reject methodological individualism, maintain that the well-being of individual human persons is the primary (maybe the sole) consideration of ethics and that from the moral point of view only individual humans who satisfy certain criteria (as stated above) can be proper subjects of moral responsibility ascriptions and objects of moral assessment and reactive attitudes such as moral indignation and resentment.

Defending the View That Institutions Can Be Moral Persons

With the publication of "The Corporation as a Moral Person" in the *American Philosophical Quarterly* in 1979, Peter A. French attacked the individualist tradition regarding institutions, organizations, and corporations by offering a functionalist conception of what it is to be a moral person and an analysis of how formal institutions can qualify as moral persons. French defended the view that corporations exhibit intentionality, are capable of rationality regarding their intentions, and are able to alter their intentions and behavior to respond to reasons, including moral reasons. He concluded that corporations and corporate-like institutions are full-fledged moral persons. Central to this argument is his depiction of the *internal decision structures* (he dubbed them CID structures) that make institutional decisions and actions possible. By coordinating, subordinating, and synthesizing the actions and intentions of individual human members of a formal organization (and often its machines), a CID structure transforms them into an institutional action taken for institutional reasons, such as promoting institutional interests. CID structures are composed of two elements: (1) an organizational chart that delineates stations and levels within the institution and (2) rules that reveal how to recognize decisions that are institutional ones and not simply personal decisions of the humans who occupy positions on the chart. Those rules are typically embedded, whether explicitly or implicitly, in corporate policy.

French's argument regarding the intentionality of institutions in the original paper of 1979 rested on his adoption of the belief/desire model of intentionality. Subsequent to criticism of his use of that model, French abandoned it and adopted Michael

Bratman's planning theory of intentionality. On Bratman's account, intentions are elements of plans—that is, deliberations and reflections prior to action. On French's revised account, a CID structure synthesizes the actions, judgments, and attitudes of individuals into the intentions (plans) and actions of an institution while also providing a mechanism for self-reflection, which is essential to its responsive function and rational decision-making processes. Institutional plans might radically diverge from those that motivate the human persons who occupy institutional positions and whose bodily movements and judgments are necessary for the institution to act.

Underlying this position is the conception of moral personhood as an artifact of *redescription.* A CID structure licenses redescribing actions of humans as the devising and executing of institutional plans, thereby revealing the institutional moral person. Gunther Teubner, along the same lines, describes an institution as an autopoietic system of actions that reproduces itself. Carlos Gomez-Jara Diez maintains that (corporate-like) institutions are not made up of human beings or even human actions. They are composed of institutional decisions and actions that construct their own social realities, which may be quite different from the reality constructions of the humans working in them.

The Debate Over Affectivity

Some who favorably view the theory that institutions are capable of intentional actions qua institutions maintain that moral personhood essentially involves *affectivity,* whereas formal institutions lack that capacity qua institutions. Institutions cannot care about the moral quality of their actions or have reactive attitudes, and so they cannot be moral persons. Those who hold the view that institutions are moral persons might argue that there are different kinds of moral personhood and institutional moral personhood does not require affectivity. Alternatively, it might be maintained that the capacity for reactive attitudes is not a necessary condition for anything to be a moral person. According to Deborah Tollefsen, what is needed is an argument to the effect that institutions do not necessarily lack affectivity, though like some humans, they may not always show it. They certainly are objects of the reactive attitudes of humans, such as resentment and indignation; on this view human expressions of reactive attitudes toward institutions are read as a demonstration that institutions are morally addressable and assessable, and that, she maintains, entails that they are presumed to have the ability to consider criticism and respond in a morally appropriate fashion. Institutions also regularly express reactive attitudes and emotions in their communications with humans and each other. Tollefsen offers a vicarious emotion theory in which employees are "conduits" for institutional emotions to account for institutional affectivity. The idea is that an institutional employee in her role qua institutional employee may be the expresser of vicarious moral emotion for the institution in its dealings with those outside the institution. Tollefsen notes that humans can have vicarious reactive attitudes and moral emotions for others even if those others do not (cannot?) have the same attitudes about themselves. So on this account, though institutions cannot directly feel moral emotions, moral emotions can be institutionalized within them.

French had maintained that institutional intentional acts are typically human bodily movements under an institutional redescription, CID structures providing epistemically transparent bases for that descriptive transformation. This thesis suggests that a similar redescription account might be applied to expose institutional affectivity, rather than depending on a vicarious emotion theory. An institution's decision structure may contain rules for the conversion of descriptions of certain types of utterances by appropriate humans into descriptions of the expression of institutional reactive attitudes.

No one claims that institutions experience "pangs" of regret or remorse or sorrow as humans might when they express reactive attitudes with regard to the behavior of others or themselves, but the ability phenomenologically to feel emotions may not be a necessary element of moral personhood. Expressions of reactive attitudes generally are performative, ritualistic, and conventional. Perhaps the expression in accord with rules is all that is required to attribute affectivity. We might, of course, regard mere ritualistic expressions as insincere if they are not "backed" by a certain sort of feeling; however, even in human affairs, an apology is not void should the apologizer not feel sorrow, regret, or remorse. If the CEO of BP expresses the corporation's regret for the Deep Water Horizon oil disaster, it seems reasonable to take as prima facie true the sentence "BP

regrets the disaster," though we typically test the sincerity of such regret by monitoring subsequent institutional behavior. Something similar seems usually to occur in the case of human expressions of emotions. Where sincerity is the issue, subsequent behavior trumps appeals to feelings. If the expression of institutional reactive attitudes and other forms of affectivity can be functionally engineered by the inclusion of rules and policies in an institution's CID structure, so that when an occupant of a certain institutional role expresses regret, sorrow, or some other emotion, that just is the institution expressing the emotion, making the apology, or regretting what it has done, then, regardless of whether or not humans in the institution, individually, collectively, or vicariously, have the appropriate emotion, institutions would seem to meet the standard conditions of moral personhood and be normatively competent, their behavior being subject to moral assessment qua institutions.

Peter A. French

See also Collective Agents; Holism, in the Social Sciences; Individualism, Methodological; Institutionalism and Institutional Theory; Reductionism in the Social Sciences; Social Institutions; Social Ontology, Recent Theories of

Further Readings

Arnold, D. G. (2006). Corporate moral agency. *Midwest Studies in Philosophy, 30,* 279–291.

Bratman, M. (2007). *Structures of agency: Essays.* Oxford, England: Oxford University Press.

Dan-Cohen, M. (1986). *Rights, persons, and organizations.* Berkeley: University of California Press.

Danley, J. (1980). Corporate moral agency: The case for anthropological bigotry. In M. Bradie & M. Brand (Eds.), *Action and responsibility: Bowling Green studies in applied philosophy* (Vol. 2, pp. 172–179). Bowling Green, OH: Bowling Green State University Press.

French, P. A. (1979). The corporation as a moral person. *American Philosophical Quarterly, 16,* 207–217.

French, P. A. (1984). *Collective and corporate responsibility.* New York, NY: Columbia University Press.

French, P. A. (1995). *Corporate ethics.* Fort Worth, TX: Harcourt Brace.

French, P. A. (2011). *War and moral dissonance.* Cambridge, England: Cambridge University Press.

Gomez-Jara Diez, C. (2008). Corporations as victims of mismanagement: Beyond the shareholders vs. managers debate. *Pace University Law Review, 28,* 101–120.

Keeley, M. (1981). Organizations as non-persons. *Journal of Value Inquiry, 15,* 149–155.

Teubner, G. (1988). Enterprise corporatism: New industrial policy and the "essence" of the legal person. *American Journal of Comparative Law, 36,* 130–155.

Tollefsen, D. (2003). Participant reactive attitudes and collective responsibility. *Philosophical Explorations, 6,* 218–234.

Tollefsen, D. (2006). The rationality of collective guilt. *Midwest Studies in Philosophy, 30,* 222–239.

Tollefsen, D. (2008). Affectivity, moral agency, and corporate-human relations. *APA Newsletters: Newsletter on Philosophy and Law, 7*(2), 9–14.

Valasquez, M. (1983). Why corporations are not morally responsible for anything they do. *Business and Professional Ethics Journal, 2,* 1–18.

Weaver, W. G. (1998). Corporations as intentional systems. *Journal of Business Ethics, 17*(1), 87–97.

Instrumentalism of Scientific Theories and Constructive Empiricism

Our scientific theories allow us to systematize and to predict observational experience; and in so doing, they frequently postulate the existence of various *unobservable entities* and processes that are supposedly responsible for, and thus help us to explain, such occurrences. For example, it is a commonplace that unsupported bodies drop; our scientific theories allow us to calculate how quickly they will fall and also tell us that this is because of the strength of the gravitational field.

An important question in the philosophy of science concerns the appropriate attitude that we should take toward this additional content. For the "scientific realist," the very fact that they allow us to successfully systematize and predict our observational experience gives us good reasons to believe that there really are microscopic organisms, gravitational fields, and atoms; and consequently, realists maintain that it is part of the scientific enterprise to investigate and to describe those parts of the external world that lie beyond our unaided senses. Ultimately, the scientific realist argues for a strong continuity between those claims of our scientific theories that concern our observational experiences and

those claims that concern the unobservable structure of the world.

By contrast, both "instrumentalists" and "constructive empiricists" have challenged this, arguing either that the unobservable content of our scientific theories does not fulfill the same semantic function as the observable content (instrumentalism) or that it does not serve the same cognitive role (constructive empiricism). Both positions are therefore committed to defending a principled distinction within our scientific theories—either logico-semantic or more broadly epistemological—and criticisms of them have largely targeted the feasibility of this project. Neither position has proved particularly popular; yet while instrumentalism has now been largely abandoned in the contemporary literature, constructive empiricism remains an ongoing research program and has evolved to include fundamental issues regarding the nature of empiricism and of the epistemology of science more generally.

Instrumentalism

According to the instrumentalist, the purpose of a scientific theory is simply to help *facilitate* the prediction of observable events. Instrumentalism is consequently committed to a sharp distinction within the language of our scientific theories, between the "observational vocabulary" that describes the observable events in which we are interested and the "theoretical vocabulary" that helps us systematize their prediction; and it maintains that the latter is to be understood as neither referring to additional unobservable events or processes nor providing any deeper level of explanation for the observable events it helps predict. On this account, a scientific theory is quite literally a *tool*—that is, an *instrument*; its theoretical vocabulary is not to be reduced or reinterpreted in terms of our observational vocabulary but shown to lack any assertoric content whatsoever.

Instrumentalism draws its inspiration from the work of the eminent German physicist Ernst Mach at the end of the 19th century, but it only achieved formal sophistication in the 1950s following the publication of various technical results by William Craig. According to Mach, science aims to provide us with an *economy of thought:* For instance, a scientific law to the effect that the acceleration due to gravity is equal to some constant does not describe the properties of some unobservable gravitational field, but it merely provides a neat summary of the results gained through observing the time taken for many individual unsupported bodies to drop to the floor. Central to Mach's position is the view that the putative content of our theoretical vocabulary is radically *dissimilar* from our observational experience—for instance, that the posited properties of atoms are clearly not mere extrapolations from the properties of the macroscopic objects that they are supposed to help explain—and thus that we do not have sufficient justification to treat our theoretical vocabulary on an ontological par with our observational vocabulary; Mach's instrumentalism is consequently a hostage to advances in our experimental technique and to the increasing justification that accrues to well-established theoretical postulates.

A more promising proposal is therefore to take the instrumentalist position as asserting a particular *logico-semantic* thesis: that, in contrast to Mach, it is not that our theoretical vocabulary is somehow too strange or insufficiently warranted but rather that it is entirely *dispensable* to the practice of science. If a scientific theory is merely a device for the systematization and prediction of observational experience, then the role of our theoretical vocabulary will be completely exhausted by the inferences it facilitates between one observable event and another. Consequently, it will in principle be possible to *eliminate* our theoretical vocabulary altogether in favor of a re-axiomatization that simply states the appropriate inferential rules entirely within the observational vocabulary of the theory. In 1953, William Craig published a result showing how this process can be made logically precise—that while the instrumentalist elimination of the theoretical vocabulary will inevitably prove to be considerably less elegant than the original scientific theory, the resulting reconstruction will be recursively enumerable and, by Craig's theorem, recursively axiomatizable.

Nevertheless, the instrumentalist re-axiomatization of a scientific theory will always consist of an infinite number of axioms, regardless of the complexity of the original theory; and, in addition, any such re-axiomatization will always be a post hoc affair, parasitic upon the existence of an already articulated scientific theory (presumably arrived at via some noninstrumentalist methodology). And while neither of these issues undermines the technical success of the instrumentalist re-axiomatization, they do indicate its epistemological undesirability.

Similarly, it is also objected that the purpose of a scientific theory is not merely to systematize and predict observable events—it is also to uncover the unobservable structure of the physical world and to explain the occurrence of observable events on their basis. These are all claims that the instrumentalist must simply deny.

More forceful objections concern the formal presuppositions of instrumentalist re-axiomatization: that it is committed to a syntactic account of scientific theories, whereby a theory is understood to consist of a set of sentences, propositions, or other linguistic devices, and that it is committed to the precise demarcation of a philosophically privileged observational vocabulary. Many contemporary philosophers of science maintain that an alternative, model-theoretic account of our scientific theories offers expressive resources consequently unavailable to the instrumentalist; and it is now widely acknowledged that the holistic nature of confirmation problematizes any attempt to cleanly differentiate between our "observational" and "theoretical" vocabularies.

Perhaps the most damaging objection to instrumentalism, however, concerns the role of probabilistic or statistical inference in the sciences. This was first formulated by Carl Hempel, who noted that instrumentalist re-axiomatization will only preserve the *deductive* inferential relations of the original scientific theory—it will therefore be completely unable to accommodate the wide range of *inductive* inferences frequently drawn in scientific practice, even those that only hold between observable events. Regardless of its philosophical merits, therefore, instrumentalism is often rejected as an inadequate account of contemporary scientific practice.

Constructive Empiricism

Constructive empiricism represents the most recent articulation of the underlying instrumentalist point of view; it has been developed—and almost exclusively defended—by Bas van Fraassen and received its first book-length presentation in 1980. Like the instrumentalist, the constructive empiricist also maintains that the aim of science is nothing more than *empirical adequacy,* that is, accuracy with respect to the observable phenomena. However, unlike the instrumentalist, the constructive empiricist does not therefore believe that our "theoretical vocabulary" can be dispensed with or indeed that the language of our scientific theories even admits of a principled distinction between its supposed subvocabularies. According to the constructive empiricist, the distinction between observable and unobservable phenomena is an *empirical* distinction, determined by the nature of the entities and processes in question rather than by the language we use to describe them; and the emphasis upon empirical adequacy is not due to a semantic inadequacy in the claims our theories make about unobservable phenomena or even due to skepticism regarding their relative justification—for the constructive empiricist, it is rather a *descriptive* claim about *scientific practice*—namely, that it just is not part of the aim of science to take us beyond observable phenomena. Consequently, while we will accept a scientific theory as a whole and continue to use its unobservable content in order to help us make predictions, construct research programs, and even offer explanations, we nevertheless *do not need to believe* what our theories say regarding atoms, fields, and other unobservable phenomena.

Constructive empiricism thus depends upon two crucial distinctions: (1) the distinction between observable and unobservable phenomena, which gives content to his notion of empirical adequacy, and (2) the distinction between *acceptance* and *belief,* which underlies how the constructive empiricist can use the unobservable content of his theories without thereby believing it.

Criticisms of the position have targeted both distinctions. With respect to the latter, it is often objected that the constructive empiricist's notion of "acceptance" is so rich that it is in fact indistinguishable from the allegedly contrasting one, "belief"—that in trying to satisfy the instrumentalist intuition that science is primarily concerned with our observational experience, yet without thereby diminishing the manifest utility of the extra-empirical content of our scientific theories, constructive empiricism simply collapses into a version of scientific realism.

Critics of the distinction between observable and unobservable phenomena have challenged both its cogency and its content. If the distinction rests upon empirical facts concerning the phenomena in question, then in order to determine its scope, we must rely upon our best scientific theories (in this case, regarding the behavior of light and the physiology of the human eye) to investigate the matter. Yet the distinction between observable and unobservable phenomena is

supposed to determine which parts of a scientific theory the constructive empiricist is to believe—including those theories that we must use in order to determine the content of the distinction. There is therefore a fundamental *circularity* in the constructive empiricist's distinction, which many take to be vicious.

Another complaint is that the constructive empiricist's distinction is *epistemologically dishonest*: that since a phenomenon counts as observable, provided there are *some* circumstances under which we would observe it (even if these circumstances could never actually be realized), there are no sound principles for privileging, say, our claims regarding the moons of a distant planet over those regarding the microscopic organisms in a petri dish. However, while this consideration may put some pressure on the constructive empiricist's contention that the aim of science is empirical adequacy, it must be remembered that the position itself does not depend upon an epistemological distinction between the claims of our scientific theories.

This last point has been made more explicit in van Fraassen's later work, where constructive empiricism is presented as part of a broader conception of empiricism and of the epistemology of the philosophy of science more generally. In van Fraassen's view, (scientific) rationality is to be thought of as a matter of *permission* rather than *obligation*—there are no universally applicable rules of inference, and an agent is justified in holding any set of beliefs provided they are not self-undermining. Specifically then, it cannot be a compelling objection to the constructive empiricist that our scientific claims regarding both observable and unobservable phenomena are on an epistemological par, since questions of justification are now inseparably bound up with questions of our epistemic *values*.

Consequently, constructive empiricism is no longer to be thought of as a rival to scientific realism but simply an alternative; and the challenge for the scientific realist, the constructive empiricist, and even the instrumentalist is merely to show that the position of each meets their own internal standards of philosophical adequacy.

Paul Dicken

See also Empiricism; Explanation, Theories of; Logical Positivism/Logical Empiricism; Metaphysics and Science; Observation and Theory-Ladenness; Realism and Anti-Realism in the Social Sciences; Scientific Method

Further Readings

Craig, W. (1956). Replacement of auxiliary assumptions. *Philosophical Review, 65,* 38–55.

Dicken, P. (2010). *Constructive empiricism: Epistemology and the philosophy of science*. Basingstoke, England: Palgrave Macmillan.

Hempel, C. G. (1958). The theoretician's dilemma: A study in the logic of theory construction. *Minnesota Studies in the Philosophy of Science, 2,* 173–226.

Mach, E. (1910). *Popular scientific lectures.* Chicago, IL: Open Court.

van Fraassen, B. C. (1980). *The scientific image.* Oxford, England: Clarendon Press.

van Fraassen, B. C. (2002). *The empirical stance.* New Haven, CT: Yale University Press.

INTELLIGENCE

Societies often differ as to which traits receive emphasis in defining intelligence: rote memory in preliterate societies, mapping among Australian Aboriginals, and so forth. On the other hand, disciplines such as philosophy or social science have also tried to pinpoint intelligence. This entry reviews various theoretical attempts at defining and measuring intelligence and raises some critical ethical points associated with such attempts.

Definitions and Theories

Different thinkers give different definitions of intelligence. Arthur Jensen thinks of it as *g* (the general intelligence factor derived from factor analysis of a variety of mental tests). This concept ranks mental skills in terms of their cognitive complexity: It places solving an arithmetic problem as more "intelligence loaded" than tasks of rote memory. Robert Sternberg believes that conventional IQ tests measure the analytic skills useful in schools and ignore creativity and practical intelligence (say, how to get people to cooperate). Howard Gardner is even more inclusive and applies the label "intelligence" to the cognitive operations of musicians and sports people.

David Wechsler, who designed the IQ tests most often used today, was more pragmatic. He used 10 subtests that collectively measure a variety of mental skills: mental acuity, or the ability to learn quickly and accurately and analyze novel situations;

information processing, or the ability to absorb information quickly about others and the world; memory and working memory, the latter referring to your ability to manipulate what is in your mind; mapping; the vocabulary we need to learn, analyze, and communicate; sufficient basic information; and learning to attack problems.

Many reject such definitions of intelligence as too imprecise. This may be a mistake, as signaled by the philosophy of science. Definitions of seminal concepts need to be general enough to accommodate a variety of proposals about *measurement*. No one would want a definition of "celestial influence" specific enough to dictate a choice between Ptolemy's sky geometry, Newton's concept of gravity, and Einstein's notion that mass affects the shape of space. No one would want a definition of the "origin of species" that dictates a choice between Darwin and creationism. Seminal concepts do not dictate which theories are viable; that is the role of theory building and evidence.

The philosophy of science does brand a certain tendency in psychology a mistake: A definition must not preempt the role of a measuring instrument. For example, Jensen once defined intelligence as what IQ tests measure. This allows no room for IQ tests to improve. No one would define heat as what the thermometers of any given time measured. It was a struggle to develop adequate thermometers, and even today, under extreme conditions, we have to develop new measuring devices.

Measurement

Given their dominance, it is worth describing the Wechsler subtests and the cognitive skills they measure: Block Design and Visual Puzzles (mental acuity); Coding and Symbol Search (information processing); Mental Arithmetic and Digit Span (working memory); Vocabulary (verbal communication); Similarities (classification); Arithmetic (numeracy); Information (acquaintance with the mechanics of the modern world); Comprehension (knowledge of everyday life).

It is often objected that such tests are biased because they measure mental traits valuable primarily in a modern industrial society. In response, it is claimed that that is what they are designed to do. However, three points are legitimate.

First, the fact that people who have not yet entered modernity and do badly on IQ tests should not be interpreted as meaning that they lack cognitive capacity. Wechsler does not assume that all societies would weigh mental skills the same. Rote memory would be more important in preliterate societies, mapping among Australian Aboriginals, and so forth. As shown by massive IQ gains over time, even our own (Western) ancestors in 1900 would have had very low IQs scored against current norms. That is because they placed less weight on analytic skills and more on the utilitarian skills they needed in everyday life. The people of developing nations are beginning to make huge IQ gains as they enter modernity.

Second, Jim Heckman and others have shown that noncognitive factors, such as motivation and self-discipline, are at least as important as IQ in predicting academic success. Third, the jury is still out on whether there are valid measures of creativity and practical intelligence or even whether or not the Wechsler tests miss these traits.

There is a debate among specialists as to the significance of massive IQ gains over time. That there have been such huge intelligence gains seems improbable. There is general agreement that human brains have no greater cognitive potential at conception than they did in 1900. There is general agreement that we are better at the analytic and classification skills that schools and professions value more today than in the past. Rather than saying our minds are more intelligent, it may make sense to say that they are more *modern*.

Ethics

Ethical problems in definitions and measurement of intelligence immediately surface: individual rights versus group membership (affirmative action policy to achieve ethnic balance), justice (as equal treatment) versus equity (equal treatment qualified by equal opportunity), utilitarian considerations (efficiency) versus individual self-esteem (employment for all), justice versus self-interest (renting your spare room to a Black male rather than waiting for, say, a Korean American female, as some do), and so on.

These problems are not solved by the slogan that we should treat everyone as an individual rather than as a member of a group. This stance can be used against affirmative action as well as against racism. Moreover, people tend to treat other people as members of a group when information about individuals is expensive. A landlord cannot afford private detectives and, therefore, uses race as a cheap information-bearing trait.

Ethical controversies in the area of intelligence and in particular over attempts to link genetics and IQ have been extremely acrimonious, as evidenced by the furor over Hans Eysenck's and others' views in the recent past. Controversies such as these often mingle philosophy and social science, blurring the distinction.

Ethical problems such as the foregoing are complicated by the question of whether ethnic differences are environmental or partly genetic. Affirmative action as a temporary expedient is different from affirmative action as a permanent option. Flynn has argued that IQ differences are primarily environmental, but other scholars who have looked at the same evidence differ. Another ethical problem is this: Should speaking the truth be qualified by its consequences? John Stuart Mill would have said that the truth cannot be racist and that the consequences of suppressing truth to spare feelings are counterproductive for everyone, including the group concerned.

James R. Flynn

See also Cognitive Sciences; Cultural Evolution; Ethical Impact of Genetic Research; Racial Critiques of Social Science Applications; Scientific Method

Further Readings

Eysenck, H. (1971). *The IQ argument: Race, intelligence and education.* New York, NY: Library Press.

Flynn, J. R. (2009). *What is intelligence? Beyond the Flynn effect.* Cambridge, England: Cambridge University Press.

Gardner, H. (1983). *Frames of mind.* New York, NY: Basic Books.

Jensen, A. R. (1998). *The G factor: The science of mental ability.* Westport, CT: Praeger.

Sternberg, R. J. (1988). *The triarchic mind: A new theory of human intelligence.* New York, NY: Penguin.

Wechsler, D. (1939). *The measurement of adult intelligence.* Baltimore, MD: Williams & Wilkins.

Intention, Social Psychology of

In our everyday lives, we use the term *intention* in the sense that our intentions are not always realized by our actions. It is this problem of *weakness of the will* or the *intention–behavior gap* that the psychology of intentions has studied. Under this conception, intentions are virtually indistinguishable from goals.

To elucidate the characteristics of such intentions or goals, this entry first discusses the weakness of the will. Next, self-regulation strategies of intention formation (goal setting) and implementation (goal striving) are introduced. Finally, the role of context for goal pursuit is highlighted—in particular, the activation of nonconscious goals and the elicitation of the "feeling of doing" or of the experience of conscious will.

Weakness of the Will

Weakness of the will has been one of the first puzzles philosophy has tried to grapple with (since the time of Socrates and Plato), but it is also central to social sciences. Weakness of the will is exhibited when agents fail to successfully pursue their intentions. A primary challenge in goal pursuit is therefore setting (committing to) goals that are not only attractive but also feasible. One method for bolstering such wise goal setting is mental contrasting of future and reality. This self-regulatory strategy asks the agent to imagine achieving a desired future outcome (e.g., getting an A in an upcoming exam) and then to imagine the most critical obstacle of reality standing in the way of achieving this future (e.g., an invitation to a party). The juxtaposition of the desired future and its obstacle automatically highlights both the perceived valence and the perceived feasibility of goal attainment. Consequently, mental contrasting strengthens commitment to and striving for goals that are perceived as not only attractive but also feasible and helps people stay away from or disengage from (attractive) goals that cannot be reached.

Goal Setting

Goals may vary not only in commitment but also in content. For instance, goals may be promotion or prevention oriented (promote good grades vs. prevent bad ones), and these facilitate goal attainment depending on whether they match the individual's self-view (ideal vs. ought) and the chosen means (eagerness vs. vigilance). Goals may contain learning versus performance outcomes (e.g., learning to solve vs. showing that one can solve problems), whereby the former type of goal fosters goal attainment when people must cope with setbacks. Finally, goals with specified standards (e.g., study for 2 hours per day) promote success more effectively than goals that keep the desired outcomes vague (e.g., study hard).

Goal Striving

Once a person has committed to a goal, it is useful to furnish it with plans specifying the where, when, and how of goal striving (i.e., form additional intentions called *implementation intentions*). It is particularly effective to lay down these plans in the format of "If I encounter situation *x*, then I will perform goal-directed behavior *y*." For example, if a student has the goal to attain an A in the upcoming test, she might form the following implementation: "If my friend invites me to her party, then I will immediately say no!" These plans derive their self-regulatory strength from the principle of strategic automaticity. The if–then structure links the selected contextual cue to an instrumental, goal-directed response, leading to increased recognition of specified cues for goal striving and the automatic initiation of the respective goal-directed response. Mental contrasting and implementation intentions complement each other in their goal-setting and goal-striving function, leading to more effective goal attainment than either self-regulation strategy separately.

Context and Unconscious Pursuit

Implementation intentions are not the only way to facilitate automatic goal striving. Cues in the agent's environment can lead to the nonconscious activation of goals. Take for example a person at a party where she does not know anyone and will never see the people there again. Even if she will walk into the party with no explicit goal to affiliate, the situational cues at the party (music, fancy clothes, etc.) will trigger her latent affiliation goal *automatically* (i.e., outside of awareness). The partygoer will thus display goal-directed behaviors such as preferring to affiliate over other tasks, continuing to socialize when interrupted, and ceasing affiliation efforts once the goal is completed. While she will not be able to report on having had this affiliation goal, one can see from her behavior that she was striving for this goal.

Finally, goals also affect people's experiences of acting. For example, goals may serve as quasi-premonitions. Privately witnessing our intentions contributes to our sense of not only having privileged access to our actions but also that we are the cause of them. Cues from the environment such as foreknowledge of the immediate event and lack of an obvious explanation of the event can trigger us to attribute an action to ourselves, though we have not *caused* the action. This phenomenon, called "the illusion of conscious will," when taken together with automatic goal striving, highlights that while intentions may be the cause of our actions, they may also be unduly influencing our beliefs regarding which actions we have caused.

Conclusion

People do not always attain their intentions or goals—a problem referred to as weakness of the will or the intention–behavior gap. The self-regulation strategy of mental contrasting fosters goal attainment by causing people to set goals that are both desirable and feasible. Goal contents can be framed in different ways (promotion vs. prevention, learning vs. performance, specific vs. vague), and the type of framing will affect the likelihood that the goal will be achieved. Difficulties on the way to achieving goals can be overcome by forming implementation intentions or if–then plans, a self-regulation strategy that guarantees goal attainment, particularly when used in combination with mental contrasting. Finally, goals may be activated outside of awareness; at the same time, awareness of goals may give agents the feeling that they caused an action they did not, in fact, effect—something that philosophical theories of introspection have identified too.

Peter M. Gollwitzer, Ana Gantman, and Gabriele Oettingen

See also Action, Philosophical Theory of; Agency; Consciousness; Goal-Directedness; Introspection (Philosophical Psychology); Unconscious Social Behavior

Further Readings

Bargh, J. A., Gollwitzer, P. M., & Oettingen, G. (2010). Motivation. In S. Fiske, D. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 268–316). New York, NY: Wiley.

Holton, R. (2009). *Willing, wanting, waiting*. New York, NY: Oxford University Press.

Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge: MIT Press.

INTENTIONALITY

Intentionality is the phenomenon of something's being *about* something or *of* something (in the sense of "of" in which a picture can be *of* something, e.g., a battle). Discussions of intentionality typically focus on *mental* intentionality—for example, thinking that the economy is in a slump, seeing the shed at the bottom of the garden, or desiring an oatmeal butterscotch cookie.

This entry focuses on contemporary philosophical accounts of intentionality and in particular on attempts to naturalize it, addresses three main obstacles such naturalizing has faced, and looks at recent approaches in this area related to cognitive phenomenology.

Naturalistic Approaches to Intentionality

Many theorists take intentionality to be an entirely natural, *physical* phenomenon, and a commitment to reductive naturalism has dominated discussion of mental intentionality in recent years. *Reductive naturalism* requires one to give an account of mental intentionality in entirely *nonmental* terms, and the consensus has been that the way to do this is to identify a natural relation that holds between states of the brain and states of the environment when and only when the former are about the latter. Philosophers give the example of tree rings "tracking" the age of trees (an entirely nonmental phenomenon) to ground the sense in which intentionality can be a natural relation. Internal states of the brain are held to track the presence of specific external conditions in a fundamentally similar way and to carry information about the environment in virtue of this tracking relation. Different theories expand on this basic idea, diverging in their more detailed expositions of the tracking relation. The dominant naturalistic theories have been causal or covariation theories and teleosemantic theories.

Problems With Naturalizing Intentionality

The "naturalization project" faces certain well-known difficulties: (1) it must account for the possibility of misrepresentation, (2) it must deal with problems of indeterminacy of content, and (3) it must respond to the worry that, once naturalized, intentionality will turn out to be ubiquitous.

1. The problem of misrepresentation is (roughly) that if it is necessary, for a mental state *S* to represent a state of affairs *X*, that *S* reliably indicate (or causally covary with) *X*, then *S* cannot occur unless *X* occurs. But this means that there is no room for misrepresentation or mistakes—that is, for a case in which *S* occurs when *X* is not the case. But clearly, a mental phenomenon can misrepresent: I can represent a dog as being black even if it is brown, or I can think that someone is Santa Claus even though Santa Claus doesn't exist.
2. Two difficulties concerning the indeterminacy of content are the "disjunction problem" (the disjunction problem may also be classified as a special case of misrepresentation) and the "stopping problem." The disjunction problem arises because a mental state *S* may systematically covary with distinct states of affairs. For example, if *S* systematically covaries with both sheep on dark nights (=*X*) and goats on dark nights (=*Y*), does *S* represent *X* or *Y*, or perhaps the disjunction [$X \vee Y$] (sheep or goats on dark nights)? The contents of our mental states do not seem indeterminate in this way.

To illustrate the "stopping problem," consider a subject, Lucy, who is perceiving or thinking about a moose called Mandy. Assuming that Lucy has the appropriate causal connections to Mandy, how does Lucy's experience or thought manage to be about Mandy rather than about the set of Mandy-caused photons impinging on her retinas, or certain other sets of causes on the causal chain leading to the thought? The antireductionist suggestion is that we need to appeal to the character of Lucy's conscious experience to explain how Lucy's experience can be said to be precisely about Mandy rather than about some other stage on the causal chain.

The disjunction problem and the stopping problem differ in that in the former the subject latches onto an object, although which one is indeterminate, whereas in the latter it is unclear how the subject latches onto an appropriate object at all.

3. If intentionality is explained in terms of "carrying information," there seems to be too much of it. The idea is that since every effect is

a reliable sign of its cause, every effect carries information about its cause, and so every effect can be said to be about or represent its cause. Fred Dretske attempts to mark off *mental* intentional states by appealing to the criterion of misrepresentation. This, however, takes us back to Problem 1.

Intentionality and Phenomenal Consciousness

In analytic philosophy of mind, the phenomenon of intentionality has shared the limelight with the phenomenon of *consciousness,* or *phenomenology.* Phenomenology can be characterized as there being *something that it is like,* experientially, to be in a mental state—as a state's having an experiential, *qualitative* character or a subjective, phenomenal character. One of the central assumptions embedded in the naturalization project has been that there is a sharp theoretical distinction between intentional properties and phenomenological properties.

It has been generally assumed that there are mental states that have intentional properties and no phenomenological properties—for example, beliefs—and mental states that have phenomenological properties and no intentional properties—for example, pain. Given the apparent independence of these two kinds of properties, it has also been assumed that if a mental state *S* has both intentional and phenomenological properties, they are logically independent of one another.

Some of these assumptions have now come under attack for both conscious perception and conscious thought. "Representationalists" or "intentionalists" have argued that the phenomenological properties of conscious perceptions supervene on or are identical to their intentional properties. More striking, though, have been some recent claims about the bearing that phenomenological properties have on the intentionality of conscious thought.

Intentionality and Cognitive Phenomenology

Until recently, it has been a standard assumption that sensory phenomenology (e.g., what it's like to see colors, taste apples, hear music) is the only kind of phenomenology there is. Now, however, there is a strong field of philosophers arguing for the existence of what is now called *cognitive phenomenology.* Cognitive phenomenology is a kind of phenomenology centrally associated with conscious thought—but also with perception and emotion—that is irreducible to any sensory phenomenology that may be associated with thought. On this view, there is something that it is like to think that 2 + 2 = 4 that is different from what it is like to think that temperance is a virtue, and this difference cannot be accounted for in terms of any sensory phenomenology (e.g., images, inner speech) that may be associated with these thoughts.

Many proponents of cognitive phenomenology hold that there is an intimate link between the cognitive-phenomenological character of thought and the intentional content of thought. There are stronger and weaker accounts of the nature of this link. According to a strong version, each thought content has a uniquely associated cognitive-phenomenological character. On this view, everyone who thinks that temperance is a virtue, for example, experiences the same cognitive-phenomenological character relative to that thought. According to a weak version, although each thought has a certain cognitive-phenomenological character, that character may vary between subjects and for a subject at different times. Many complications arise in the detailed spelling out of this view, one of which has to do with whether one thinks that the kind of content associated with thought is broad or narrow, or a combination of these.

Philosophers have appealed to cognitive phenomenology to solve all three of the main outstanding problems faced by the naturalization project (Points 1–3 above). Uriah Kriegel argues for a partial solution to the misrepresentation problem by claiming that thinking about nonexistent objects is a matter of instantiating the relevant phenomenal character. Terence Horgan and John Tienson suggest that the determinacy of thought is partly grounded in phenomenal character. Given that what it is like to think about sheep is distinct from what it is like to think about goats on dark nights, we can appeal to this difference in "what it's likeness" to explain why an intentional state has one content rather than another. Galen Strawson responds to the stopping problem by claiming that the intentionality of Lucy's thought can zero in on Mandy the moose rather than any other causes along the causal chain because Lucy's experience includes a cognitive-phenomenological element: her consciously taking her experience to be an experience of a physical object at a certain distance from

her (rather than, say, an experience of photons). (See Jerry Fodor's 2008 book for a nonphenomenological solution to the stopping problem.)

Michelle Montague

See also Cognitive Phenomenology; Collective Intentionality; Consciousness; Introspection (Philosophical Psychology); Mind–Body Relation; Reduction and the Unity of Science; Supervenience; Teleosemantics

Further Readings

Bayne, T., & Montague, M. (Eds.). (2011). *Cognitive phenomenology.* Oxford, England: Oxford University Press.

Chalmers, D. (2004). The representational character of experience. In B. Leiter (Ed.), *The future of philosophy* (pp. 153–181). Oxford, England: Oxford University Press.

Dretske, F. (1981). *Knowledge and the flow of information.* Cambridge: MIT Press.

Dretske, F. (1995). *Naturalizing the mind.* Cambridge: MIT Press.

Fodor, J. (1982). *Psychosemantics.* Cambridge: MIT Press.

Fodor, J. (2008). *LOT 2: The language of thought revisited.* Oxford, England: Oxford University Press.

Harman, G. (1990). The intrinsic quality of experience. *Philosophical Perspectives, 4,* 31–52.

Horgan, T., & Tienson, J. (2002). The intentionality of phenomenology and the phenomenology of intentionality. In D. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 520–533). Oxford, England: Oxford University Press.

Kriegel, U. (2007). Intentional inexistence and phenomenal intentionality. *Philosophical Perspectives, 21,* 307–340.

Lyons, W. (1995). *Approaches to intentionality.* Oxford, England: Oxford University Press.

Millikan, R. (1984). *Language, thought, and other biological categories.* Cambridge: MIT Press.

Montague, M. (2010). Recent work on intentionality. *Analysis, 70*(4), 765–782.

Papineau, D. (1993). *Philosophical naturalism.* Oxford, England: Blackwell.

Pitt, D. (2004). The phenomenology of cognition, or, what is it like to think that P? *Philosophy and Phenomenological Research, 69*(1), 1–36.

Siewert, C. (2011). Phenomenal thought. In T. Bayne & M. Montague (Eds.), *Cognitive phenomenology* (pp. 236–267). Oxford, England: Oxford University Press.

Strawson, G. (2008). Real intentionality 3: Why intentionality entails consciousness. In *Real materialism and other essays* (pp. 281–306). Oxford, England: Oxford University Press.

Strawson, G. (2011). Cognitive phenomenology: Real life. In T. Bayne & M. Montague (Eds.), *Cognitive phenomenology* (pp. 285–325). Oxford, England: Oxford University Press.

INTERDISCIPLINARITY

This entry presents central features as well as different meanings of interdisciplinarity and its importance for the social sciences.

In the modern era, concepts pass in and out of fashion quickly. But some new coinages stick: They fit new and enduring needs. *Interdisciplinarity* is a term that seems destined to stay, for the term defines a new era of knowledge, the chief characteristics of which are the need to address problems that resist parameterizing, an insistence on a tight link between knowledge production and its use, and the democratizing of knowledge production.

Disciplines, a Western invention, did not truly come into existence until the second half of the 19th century. Until that time, colleges and universities emphasized the perennial aspects of knowledge, and the importance of passing down received wisdom from generation to generation. Mid-19th-century urbanization and industrialization changed this: New demands were made on institutions of knowledge to tie their work to the growing capitalist economy. Majors were created and disciplines formed, and professors were now expected to produce new knowledge rather than simply pass down received wisdom. This led to the development of disciplinary expertise and the progressive disaggregation of knowledge.

Calls for "interdisciplinarity" were the expected result. The term first appeared in Social Science Research Council awards listings and reports beginning in 1930. Since the 1960s, the term has become a commonplace across the academy—ubiquitous in strategic plans, curricular innovations, and research projects. There is a tendency, however, for the term to be treated as a code word for "relevance" rather than as an operational term with distinct meanings.

Meanings of Interdisciplinarity

First and most simply, interdisciplinarity is one of a set of cognate terms that describe different ways to organize knowledge. A *discipline* is a body of knowledge thought to be more or less discrete from other regions of knowledge. A *multidisciplinary* approach to knowledge seeks to link or juxtapose disparate bodies of knowledge without essentially changing those bodies of knowledge. In contrast, an interdisciplinary approach to knowledge emphasizes not the simple juxtaposition but rather the blending or integration of different types of knowledge.

All of these terms, however, designate research that stays within the walls of the academy. The term *transdisciplinarity* has been used in a number of ways—for instance, it has been used to describe the goal of unifying all of knowledge under a master trope or concept such as systems theory. But its predominant meaning today refers to efforts to adapt academic knowledge production for problem solving out in the larger world. Note that work can be transdisciplinary without being interdisciplinary in nature: It is possible for the knowledge of a single discipline to be used in a transdisciplinary project. Nor should these different types of knowledge be treated as hierarchical. *Inter-* or *trans*disciplinary knowledge is not "better" or more highly developed than multidisciplinary knowledge, except in terms of a specific goal or end.

These definitions are well established today—although it is worth noting that *interdisciplinarity* is also used as a more general term, including the meaning of transdisciplinarity—especially in the United States. But there is significant scholarly disagreement today about the overall goals or purpose of studying interdisciplinarity.

For some, especially those focused on the undergraduate experience, concerns with methodology dominate. Some researchers seek to identify a set of interdisciplinary procedures for the efficient development of undergraduate programs in areas such as environmental studies and Chicano studies. More problematically, they also seek to identify principles and methodology for interdisciplinary research. Others take a more descriptive approach. Some interdisciplinarians engage in what is sometimes called the "science of team science." For instance, members of different disciplines who share a research project may come together in a workshop setting under the auspices of a researcher, where they take a survey that helps them become more self-aware of their distinctive methodological and philosophical assumptions.

One can also find various types and degrees of *antimethodologism* within the study of interdisciplinarity. Some argue that interdisciplinarity is more a matter of manner than of method. While granting that those experienced in interdisciplinary education and research can identify a set of rules of thumb, interdisciplinary work is seen as primarily involving the cultivation of a set of personal virtues such as open-mindedness, disciplinary modesty, and the ability to see things from different perspectives. For instance, Wolfgang Krohn highlights the importance of a case-based approach to understanding interdisciplinarity. He views interdisciplinary research as involving a distinctive relationship between individual cases and more general knowledge. Rather than seeking lawlike principles to form part of a method, interdisciplinary knowledge calls for a critical reassessment of our concept of scientific laws and general principles.

Third, interdisciplinarity is also taken as highlighting an ongoing shift in culture. In this interpretation, *interdisciplinarity* points up increased attention to understanding the relation between knowledge production and its use. In its classic 19th- and 20th-century form, disciplinary knowledge assumed that knowledge was both inherently relevant and fundamentally benign. No extra step was necessary to make knowledge relevant to society. And if something bad resulted from the use of knowledge, this was not considered the responsibility of the producer of that knowledge. Today, however, it is becoming increasingly evident that knowledge must be *coproduced*—a process where the users of knowledge are involved in the very design and production of that knowledge. One sign of this coproduction is the development of user-driven content on the Internet via Web 2.0.

Conclusion

In closing, it is worth noting that the study of interdisciplinarity has become more professional in recent years. One finds a significant scholarly literature, including compendia such as the *Oxford Handbook of Interdisciplinarity.* There has also been a notable degree of institutionalization of inter- and transdisciplinarity, via centers, institutes, meetings, and associations. In Europe, TDnet has been

running an annual conference since 2004, and in the United States, the Association for Integrative Studies has been holding an annual meeting since 1979. In 2008, the Center for the Study for Interdisciplinarity was created at the University of North Texas. In 2011, these three groups joined to create INIT, the International Network for Interdisciplinarity and Transdisciplinarity, in Utrecht, The Netherlands. Given the increasing societal demands for greater relevance to academic research, we can expect the focus on interdisciplinarity to increase in the future.

Robert Frodeman

See also Disciplinarity; Knowledge Society; Paradigms of Social Science; Postmodernism; Social Epistemology

Further Readings

Frodeman, R., Klein, J. T., & Mitcham, C. (Eds.). (2010). *The Oxford handbook of interdisciplinarity.* Oxford, England: Oxford University Press.

Hadorn, G. H., Hoffman-Riem, H., Biber-Klemm, S., Grossenbacherp-Mansuy, W., & Joye, D. (2008). *Handbook of transdisciplinary research*. New York, NY: Springer.

Klein, J. T. (2000). *Creating interdisciplinary campus cultures: A model for strength and sustainability.* San Francisco, CA: Jossey-Bass.

Krohn, W. (2010). Interdisciplinary cases and disciplinary knowledge. In R. Frodeman (Ed.), *The Oxford handbook of interdisciplinarity* (pp. 31–38). Oxford, England: Oxford University Press.

Nowotny, H., Scott, P., & Gibbons, M. (2003). *Re-thinking science: Knowledge and the public in an age of uncertainty.* Cambridge, England: Polity Press.

Stehr, N., & Weingart, P. (2000). *Practising interdisciplinarity.* Toronto, Ontario, Canada: University of Toronto Press.

International Relations, Philosophical and Methodological Debates

International relations emerged as a self-conscious discipline only at the beginning of the 20th century and by its end was in danger of vanishing in an amorphous field called *global studies*. The question is whether a discipline premised on the independence of the territorial state can survive the decline of that independence. Another question is whether its subject matter—interstate relations—poses a unique challenge to inquiry or whether its main challenges are those it shares with other human sciences. These include the tension between scientific and nonscientific modes of understanding and between the methods of inquiry they imply.

The study of international relations has thus raised crucial methodological and ontological questions of interest to the philosophy of the social sciences (questions of explanation or regarding the status of collective entities such as the state, issues of agency, etc.). In addition, the field of international relations has raised a number of important moral questions (e.g., just-war theories). Therefore, philosophy has been intimately linked with this social science. This entry considers whether there is a distinct international relations discipline, surveys debates over whether the discipline is a "scientific" one, and examines the relationship between the knowledge it offers and the methods needed to generate that knowledge.

Disciplinary Identity

The state is an idea, not a thing. In Western thought, it began as a legal idea in arguments about the claims of rulers to authority within a realm against internal rivals and external claimants (e.g., the papacy and the Holy Roman Empire). A state came to be understood as a community of persons united under a common body of law emanating from an authority recognized to be *sovereign*—that is, superior within its boundaries to any other authority, internal or external.

This legal debate was paralleled by debates about the moral foundation of such authority and its effectiveness. Considering states in relation to one another generates distinct questions, which we recognize as questions of international law, international ethics, and foreign policy or diplomacy, understood as the acquisition, preservation, and use of state power. It also raises philosophical questions, such as whether states can be said to have motives or make decisions. Scholars have investigated all these questions, and the discipline of international relations has been thus identified at times as a branch of jurisprudence or political philosophy and at other

times as an autonomous discipline concerned with power politics.

Disciplinary autonomy premised on the study of power is precarious, however, because power is not a quality unique to states. What distinguishes international relations is not that states are powers (many are not) but that they are, by definition, legally independent territorial associations that have relations with one another. But other entities including individuals, companies, nongovernmental associations, and criminal gangs also engage in transactions across territorial boundaries. If we consider these transactions, limiting the discipline to relations between governments fetishizes the state as the "thing" that defines the field, making a claim that is best defended as a choice justified by its results. The question of whether the results are worth the cost is what divides political realists from those who think that studying other kinds of cross-national relations is also worthwhile.

If national boundaries are losing significance, cross-national relations will cease to be distinctive. For those who see this happening, international relations is morphing into global relations, as the phrases of the moment—global warming, global migration, global democracy, global justice—suggest. Although the adjective "international" adds meaning to words like *migration* or *justice,* "global" does not. There is no clear distinction between the idea of justice, unmodified, and global justice. The emerging discipline of global studies—now a popular major in many universities—will face an identity crisis worse than that experienced by the discipline of international relations. Perhaps a sensible conclusion would be one that recognized that international relations can be studied in different ways and that to study it does not require a distinct and autonomous academic discipline.

The Science Question

Since the middle of the 20th century, the study of international relations has been embroiled in a debate about the degree to which that study could be scientific. Initially, the model of science was physics, which (following philosophers in the tradition of the late-19th-century and early-20th-century positivism) was believed to consist of general statements of invariant regularity ("covering laws") that could be combined with factual data to explain and predict particular events. This view of scientific explanation was adopted by the movement known as *behaviorism* (in political science, *behavioralism*), with psychology as a more proximate model. In the last decades of the 20th century, economics became the field to imitate. At the same time, a series of counterpositions emerged, first *traditionalism* or the classical approach and then *postpositivism,* a blanket label covering a variety of approaches, some rejecting and others claiming the mantle of science.

Some perspective on these disagreements can be gained by considering the German origins of political science, which in turn influenced the study of international politics. In 19th-century German thought, history had as much prestige as science. With the models of physics and history in mind, scholars at the end of that century distinguished between the natural and the human sciences, the latter (*Geisteswissenschaften*) defined by their concern not only with particulars but with meaning and therefore with interpretation. "Science," in this context, denoted the kind of inquiry appropriate to understanding and explaining intelligent human conduct, in contrast to inquiry into not-intelligent (natural) processes. Science, in short, was nothing more than systematic research governed by the canons of objectivity and validation appropriate to what was being studied.

This history is often forgotten by those who argue that the study of international relations should be scientific in the same way as the natural sciences. One response to that claim is to use the definition of science as systematic research and to advocate for a pluralist science of international relations. Another is to view "science" as one among a number of alternative modes of understanding and inquiry. Science, here, is natural science—the study of not-intelligent processes. History and other humanistic inquiries, in contrast, are those that interpret and explain intelligent human conduct in terms of ideas and meaning. And both are distinct from activities involved in making moral, legal, or political decisions, which are practical rather than explanatory activities.

To privilege any one of these modes in making sense of international relations is to mistake a mode of knowledge, which is by definition limited and partial, for the whole of knowledge. Thus, scientism is the error of privileging the natural sciences as the model of all genuine knowledge. Historicism is the error of thinking that nothing can be understood

except in relation to time and place. And pragmatism, which includes many versions of Marxism and critical theory, is the error of confusing theoretical inquiry with practical action. If theory is always for someone and for some purpose, its objectivity is always suspect, and the claim itself is exposed as self-refuting.

The failure to distinguish distinct modes of understanding and inquiry pervades the study of international relations, as it does other social sciences. Discussions of causal explanation, for example, often assume—mistakenly—that the idea of cause in physics or biology is the same as in history or jurisprudence. We can generalize about the causes of war, but such generalizations may not tell us much about the causes of this or that war or about who should be held responsible for causing it. To explain a decision that involves thought and ideas, one must know the antecedents that make it intelligible. Also fatal to explanation is the error of combining distinct modes, as if arguments in one mode could support or deny arguments in another. This error is found in realist and Marxist efforts to extract policy advice from presumed necessities and constructivist efforts to combine interpretive and naturalistic considerations within a single theory.

Recognizing that making sense of international relations benefits from different kinds of inquiry implies a pluralist understanding of the field. Whether its practitioners embrace pluralism by adopting a broad definition of science as systematic scholarship or define science narrowly as one mode of scholarship among others, they move beyond the barbarism of insisting that their preferred mode dominate the others. The humility gained through this recognition of modal pluralism can be used to good effect in appreciating what can and cannot be explained within the limits of a given mode and weeding out theories that fail the test of modal coherence.

Modes and Methods

If there are different modes of understanding and inquiry, it follows that research will require methods appropriate to the kinds of understanding possible in each. Modal pluralism implies methodological pluralism. It is widely acknowledged that we should start with questions rather than with methods, but the practice of teaching research using teaching methods undercuts this precept. The problem is made worse when quantitative methods are privileged. Doing so reinforces the assumption that all research is empirical and that the so-called qualitative methods are merely an alternative way of generating and analyzing empirical data and making empirical generalizations and causal inferences. All research from this point of view is "scientific" research. Antiscientistic epistemologies challenge this assumption and invite different methods for generating knowledge.

The student of diplomacy, for example, might need the skills of an anthropologist to understand contemporary diplomatic practices and those of a historian to understand the practices of earlier periods. Depending on their research questions, students of diplomatic discourse might use the methods of quantitative content analysis or rhetorical classification. Scholars who think the ideas of the people they are studying are important in explaining the behavior of those people will want information that differs from data collected by scholars who dismiss such self-understandings, and they will use different methods to obtain and analyze that information. Someone with a practical interest in just-war thinking might use its conventional principles to make judgments about a particular war; in contrast, someone more philosophically inclined might abstract a self-consistent theory of just war from arguments about particular cases, generating principles that differ significantly from the conventional ones. One might accept the principle that in combat enemies can be killed without blame, while another might reject that principle and in doing so challenge the received distinction between war and murder.

Social scientists need to make assumptions to get on with their research, but they also need to question assumptions to do innovative research. Thinking carefully about one's assumptions is logical prior to designing one's research, and laying out the research design must come before the choice of methods. An intelligent approach to research must start by recognizing the plural character of knowledge and proceed from clarity about the kind of knowledge sought to choosing designs and methods appropriate to generating it.

What are commonly identified as theories of international relations—realism, liberalism, constructivism, critical theory, poststructuralism, the English school, and others—are neither distinct

modes of understanding nor alternative methods of inquiry. Each provides a vocabulary for debating philosophical and methodological questions. Unfortunately, those vocabularies often carry one effortlessly to conclusions that are already implicit in them. Theory is best seen as an activity—the activity of theorizing—not as a body of conclusions. One can theorize by working out the implications of a set of assumptions, but one can also theorize by questioning those assumptions and heading off in a new direction. In doing so, theorists must rely on their own experience and insight and not look to philosophy for guidance, as naive scholars do when they invoke "the philosophy of science" as authoritative for scholarship. In this respect, the study of international relations is no different from any other study.

Terry Nardin

See also Behavioralism in Political Science; Covering-Law Model; Explanation Versus Understanding; Institutions as Moral Persons; *Naturwissenschaften* Versus *Geisteswissenschaften;* Normativism Versus Realism; Realism and Anti-Realism in the Social Sciences

Further Readings

Brown, C., Nardin, T., & Rengger, N. (Eds.). (2002). *International relations in political thought: Texts from the Ancient Greeks to the First World War.* Cambridge, England: Cambridge University Press.

Carlsnaes, W., Risse, T., & Simmons, B. A. (Eds.). (2002). *Handbook of international relations.* London, England: Sage.

Dunne, T., Kurki, M., & Smith, S. (Eds.). (2010). *International relations theories: Discipline and diversity* (2nd ed.). Oxford, England: Oxford University Press.

Hollis, M., & Smith, S. (1990). *Explaining and understanding international relations.* Oxford, England: Oxford University Press.

Jackson, P. T. (2011). *The conduct of inquiry in international relations.* New York, NY: Routledge.

Mapel, D., & Nardin, T. (Eds.). (1998). *International society: Diverse ethical perspectives.* Princeton, NJ: Princeton University Press.

Nardin, T. (2009). International political theory. In S. Burchill, A. Linklater, R. Devetak, J. Donnelly, T. Nardin, M. Paterson, . . . J. True (Eds.), *Theories of international relations* (4th ed., pp. 284–310). Basingstoke, England: Palgrave Macmillan.

Reuss-Smit, C., & Snidal, D. (2008). *The Oxford handbook of international relations.* Oxford, England: Oxford University Press.

Schmidt, B. C. (1998). *The political discourse of anarchy: A disciplinary history of international relations.* Albany: SUNY Press.

INTERSUBJECTIVITY

Intersubjectivity is an interdisciplinary concept that refers to the field of interaction between the self and other. Intersubjectivity has its origins in the philosophical rejection of René Descartes's philosophy of consciousness (as a private datum) and his notion of the isolated mind. From an intersubjective perspective, human beings exist not in isolation but in a world with others. The concept of intersubjectivity is of primary interest to philosophers, social and political theorists, and psychoanalytic psychologists. This entry provides an account of this notion in these fields.

Philosophy

From a philosophical perspective, intersubjectivity is generally associated with the continental tradition rather than the analytic and postanalytic tradition. However, parallels can be found with Ludwig Wittgenstein's philosophy of language, according to which humans are inherently connected through their use of language. Intersubjectivity is one of the dominant themes in the tradition of existential phenomenology, though not all thinkers who write about intersubjectivity employ the term.

In Edmund Husserl's phenomenology, intersubjectivity refers to the process by which other human beings are experienced as subjects rather than as objects in a separate world. From the perspective of phenomenology, Husserl emphasized that the world we intend is not a private world but a world common to and accessible to all. The lingering monadism at work in Husserl's philosophy led his student Martin Heidegger to reject the term *intersubjectivity.* Instead, Heidegger coined the notions of *Dasein* and *being-in-the-world* to refer to the situated nature of the human being and its fundamental contextualism. For Heidegger, the human being cannot be separated from the social contexts in which

experience always and already unfolds. However, Heidegger's account of sociality remains peculiarly underdeveloped and has been criticized for its ethical and political shortcomings by the contemporary continental philosophers Emmanuel Levinas and Jürgen Habermas, respectively.

A different perspective on intersubjectivity was introduced by Martin Buber, who delineates two modes of social being: *I–It* and *I–Thou*. For Buber, the subject only exists insofar as it swings "between" these two types of relations. The French philosophers Maurice Merleau-Ponty and Jean-Paul Sartre (both inspired by phenomenology, the latter being one of the fathers of existentialism) moved beyond the perspectives of Husserl and Heidegger to examine the role of embodiment and recognition in the social sphere. For Merleau-Ponty, the perception of others and the world occurs through the body, thus illustrating the fundamental embodiment of all human experience. For Sartre, social interaction is delineated in the "dialectic of the look," according to which the other is an immediate and threatening presence that I must seek to subjugate in turn. Sartre's account of the struggle for recognition is similar to Alexander Kojève's influential reading of modes of social interaction in G. W. F. Hegel's philosophy.

Social Theory

For social theorists, intersubjectivity provides a means to overcome the impasses associated with the philosophy of consciousness and a subject-centered notion of reason. Habermas uses the concept of intersubjectivity to develop an emancipatory social philosophy. Drawing on George Herbert Mead's account of symbolic interaction, Habermas's communicative framework is based on a symbolically structured life-world, in which subjectivity is constituted through reflexive linguistic interaction. According to Habermas, the potential for communicative reason is embedded in language. Based on the assumption that recognition of truth can be attained through dialogue, Habermas postulates a counterfactual "ideal speech situation" in which genuine communication and mutual recognition can be achieved. Habermas thus takes linguistic intersubjectivity as his starting point and construes subjectivity in relation to it.

A different approach to the concept of intersubjectivity is proposed by Anthony Giddens. His theory of *structuration* suggests that the reflective capabilities of the subject are inherently linked to social experience. The concepts of subject and society are not simply interacting but fundamentally interdependent. Neither can be understood or conceptualized without the other, and neither can exist by virtue of a recursive process. Giddens refers to the coproduction of action and structure as the recursive nature of social life, in which the structured properties of social activity are constantly re-created out of the very resources that constitute them. Human agency, on this view, is neither a capacity of the individual nor a function of the social context. Because human beings act in and through structures and systems, we are not independent actors in the world, and our actions are never entirely of our own choosing. In contrast to Habermas, for whom the subject is linguistically constituted, Giddens underlines the reflexive capacity of the subject but avoids the strict association of the subject with rational reflection and language.

Psychology and Psychoanalysis

For psychologists and psychoanalysts, the concept of intersubjectivity is interwoven with emotion, imagination, and embodiment. Intersubjectivity has become an important theme used by Daniel Stern to describe a progression of developmental stages, by Jessica Benjamin to define a feminist developmental theory, and by Robert Stolorow, George Atwood, and Donna Orange to refer to a mode of psychoanalytic theory and practice. Intersubjectivity was introduced to psychoanalytic psychology by Ludwig Binswanger and Jacques Lacan, who attended to the intersubjective nature of human experience, as evidenced in the patient's interpersonal and linguistic contexts.

In contemporary psychoanalysis, intersubjectivity refers to the ongoing and reciprocal engagement of patient and therapist in a mutually transforming interaction. In contrast to the traditional Freudian emphasis on the intrapsychic experience of the patient and the neutrality of the therapist, the intersubjective perspective focuses on the interacting subjective worlds of patient and therapist. The therapeutic relationship is seen as the domain of analytic inquiry, and the goal of treatment is to reach an understanding of psychological phenomena as they emerge between the patient and the therapist.

The intersubjective therapeutic perspective has evolved into a broad-based philosophy of practice termed *contextualism*. A common critique of intersubjectivity is that it fails sufficiently to account for the broader contexts of experience and remains focused on the individual subjectivities of patient and therapist. In contrast to traditional philosophical accounts of intersubjectivity, contextualism seeks to account for the sociocultural constitution of all experience. On this view, the psychological processes are actually made up within, as opposed to merely facilitated by, culture and society.

Roger Frie

See also Being-in-the-World; Cognitive Phenomenology; Consciousness; Ego; Existential Phenomenology and the Social Sciences; Self and the Social Sciences; Self-Knowledge

Further Readings

Crossley, N. (1996). *Intersubjectivity: The fabric of social becoming*. London, England: Sage.

Frie, R. (1997). *Subjectivity and intersubjectivity in philosophy and psychoanalysis*. Lanham, MD: Rowman & Littlefield.

Introspection (Philosophical Psychology)

Etymologically, introspection means "looking within." The term *introspection* is typically used as a placeholder for the distinctive way, whatever it is, in which we know our own minds. Yet it is contentious whether we know our own minds by inner perception. The task for a theory of introspection is to provide a more substantive account of the way in which we know our own minds.

This entry surveys various theories of introspection, including inner sense theories, inferential theories, causal theories, constitutivist theories, transparency theories, and acquaintance theories. The final section raises further questions about the scope of introspection, its reliability, and its theoretical implications for wider issues in philosophy and cognitive science.

Inner Sense Theories

A classical view associated with modern European philosophy, and in particular with René Descartes, John Locke, and David Hume, is that introspection is a form of inner sense—that is, inner perception of one's own mind. On this view, there is an asymmetry between first-person and third-person perspectives such that the way we know our own minds is both *different* from and *better* than the way we know the minds of others. It is different because we know our own minds directly on the basis of inner perception, whereas we know the minds of others indirectly on the basis of inference from external perception of physical behavior. And it is better because inner perception, unlike external perception, is claimed to be immune from ignorance and error.

The problem with this classical view is that the notion of inner sense remains obscure. Does inner sense involve a sensory organ, such as an inner eye? And if not, then what is the substance of the claim that introspection is a form of nonsensory perception?

Inferential Theories

The 20th-century English philosopher Gilbert Ryle famously derides "Descartes's myth" of inner sense as a hopelessly confused metaphor. Ryle claims instead that the way in which we know our own minds is by inference from observation of our own physical behavior. On Ryle's theory, the way we know our own minds is no different from the way we know the minds of others; moreover, it is no better except insofar as we have more observational data about our own behavior than about others. Hence, there is no relevant asymmetry between first-person and third-person perspectives.

The problem with Ryle's theory is that it cannot explain the introspective self-knowledge that we plausibly have. For instance, I know that I am currently thinking about tomatoes, but I cannot know this by inference from perception of my behavior, since my current thoughts about tomatoes exert no causal influence on my behavior.

Ryle's inferentialism is motivated by his *behaviorism*: If the metaphysics of mind consist of behavioral dispositions, then the epistemology of mind consists of knowledge of those dispositions. However, others have attempted

to divorce inferentialism from these behaviorist commitments. For instance, the psychologist Alison Gopnik argues on the basis of evidence from developmental psychology that knowledge of one's intentional states relies on unconscious inferences informed by the acquisition of a theory of mind. However, she claims that these inferences rely not only on premises about behavior but also on conscious experience—what she calls "the Cartesian buzz." Thus, her inferentialism is restricted to knowledge of intentionality rather than conscious experience. Similarly, the philosopher Peter Carruthers defends an inferentialist theory restricted to our knowledge of cognition rather than perception.

Causal Theories

Ryle's behaviorism is widely rejected in favor of *functionalism,* which defines mental states by their causal roles in the production of other mental states as well as physical behavior. The philosopher David Armstrong, an early pioneer of functionalism, provides a causal analysis of the metaphor of introspection as inner perception. Perception, according to Armstrong, is merely the acquisition of true or false beliefs about the world by means of an appropriate causal mechanism. Accordingly, introspection involves the operation of a "self-scanning mechanism" that yields true or false beliefs about one's own mental states. Armstrong argues against the Cartesian view that introspection is immune from ignorance and error, on the grounds that no causal mechanism is immune to malfunction. Moreover, he argues that the reliability of our introspective mechanisms is sufficient to explain the possibility of introspective knowledge.

The contemporary philosophical literature contains various analogs of Armstrong's self-scanning mechanism. For instance, Shaun Nichols and Stephen Stich posit a "monitoring mechanism" that takes a representation that *p* as input and yields a representation that I believe that *p* as output, together with a related cluster of perceptual monitoring mechanisms designed to explain our introspective knowledge of perceptual states. Similarly, Alvin Goldman proposes a causal theory on which introspective mechanisms are sensitive to the neural states that realize mental states.

Constitutivist Theories

The philosopher Sydney Shoemaker classifies causal theories as examples of a "broad perceptual model" of introspection, which he rejects in favor of a *constitutivist* theory of introspection. According to causal theories, our introspective beliefs are independent of and caused by the mental states they are about, whereas according to constitutive theories, our introspective beliefs are partially constituted by those mental states.

Shoemaker argues for constitutivism on the grounds that there is no genuine possibility of *self-blindness* in which one is rational and one has concepts of one's mental states but one has no introspective self-knowledge of one's mental states. According to Shoemaker, any rational and conceptually competent subject who is in a mental state of the relevant kind thereby believes that she is in that mental state. Thus, he argues, introspective beliefs are partially constituted by the mental states they are about together with certain background conditions such as rationality and conceptual competence. On this view, introspective beliefs are neither independent of nor caused by those mental states but are partially constituted by them.

Shoemaker's arguments against the possibility of self-blindness have received criticism, but many critics acknowledge a close relationship between rationality and introspective self-knowledge. Intuitively, rationality is consistent with massive ignorance and error about the external world but not one's own mental states. Therefore, the challenge for causal theories of introspection is either to explain the connection between rationality and self-knowledge or to explain it away.

Transparency Theories

Others argue against inner sense theories on the grounds that introspection is a matter of "looking outward" rather than "looking within." For instance, Gareth Evans observes that if I want to know whether I believe there will be a third world war, I can simply ask myself, "Will there be a third world war?" Evans claims that belief is *transparent* in the sense that I can answer the question whether I believe that *p* by answering the question whether *p*. Moreover, if I form the second-order belief that I believe that *p* on the basis of my first-order belief

that *p,* then my second-order belief is guaranteed to be true, which promises to vindicate the Cartesian commitment to the infallibility of introspection.

It is an open question whether transparency extends from belief to other mental states. Evans argues that perceptual experience is transparent in the sense that I can answer the question whether it perceptually seems to me that *p* by answering the question whether *p* on the basis of the way things perceptually seem, while excluding any extraneous information. However, this threatens the vindication of Cartesian infallibility, since mistaken perceptual beliefs about the world can lead to mistaken introspective beliefs about how things seem perceptually. Recent work by Alex Byrne explores analogs of transparency for other mental states, including desire, intention, and thought.

The phenomenon of transparency gives rise to a puzzle. What justifies me in treating evidence about the world as evidence of my own mental states? Richard Moran argues that the transparency phenomenon reflects a fact about rationality—namely, that I am rationally entitled to take my beliefs to reflect my evidence about the world. Here again, the connection between rationality and self-knowledge is a crucial theme.

Acquaintance Theories

Proponents of acquaintance theories, including David Chalmers, argue that introspective knowledge of conscious experience is made possible by *acquaintance*—that is, a relation to conscious experience that explains one's ability to acquire introspective knowledge of conscious experience. Chalmers defines acquaintance by its epistemic role, but this raises the question "What is the nature of the acquaintance relation that plays this epistemic role?"

One answer is suggested by Terry Horgan and Uriah Kriegel's claim that conscious experience is *self-representing* in the sense that conscious experience represents itself. Moreover, Horgan and Kriegel argue that the self-representation built into conscious experience constitutes a form of acquaintance with conscious experience and thereby explains the possibility of introspective knowledge. However, it is debatable whether we need the claim that conscious experience is self-representing in order to explain our introspective knowledge of conscious experience.

Further Questions

It seems obvious that each of us has a capacity for introspective knowledge of our own minds, which is not available to anyone else. However, it is much less obvious how a theory of introspection should explain this capacity for introspective knowledge. Moreover, in addition to questions about the nature of introspection, further questions arise concerning its scope, reliability, and theoretical significance:

On the scope of introspection: Which mental states can we know by introspection? On a classical Cartesian conception of mind, we know all our mental states by introspection. However, the Freudian and Chomskian revolutions have shown that there are unconscious mental states that cannot be known by introspection. Therefore, some restrict the scope of introspection to one's current stream of conscious experience, while others claim that one's introspective knowledge extends to one's standing beliefs and desires.

On the reliability of introspection: Is there any sense in which introspection is more reliable than other ways of knowing about the world? On the classical Cartesian conception, introspection is immune from ignorance and error. However, there is psychological evidence that we are often mistaken about our mental states and processes—including the classic experiments by Richard Nisbett and Timothy Wilson in which subjects confabulate the reasons for action, such as choosing a pair of socks. One response is to argue that introspection is immune from ignorance and error when it concerns one's current stream of conscious experience. However, Eric Schwitzgebel has recently argued that introspection of conscious experience is even less reliable than sensory perception.

On the theoretical significance of introspection: What are the implications of these questions about introspection for wider issues in philosophy and cognitive science? First, there are important questions in epistemology about whether introspective knowledge is immune from skeptical doubt, whether it serves as a foundation of empirical knowledge, and whether it provides us with reflective access to our reasons for belief and action. Second, there are important questions in philosophy of mind and cognitive science about whether introspection is sufficiently reliable to

provide a source of evidence about conscious experience and its neural correlates. Therefore, questions about the nature, scope, and reliability of introspection have important consequences for central issues in philosophy and cognitive science.

Declan Smithies

See also Behaviorism, Philosophical Conception of; Consciousness; Intention, Social Psychology of; Intentionality; Metacognition and Agency; Self and Essential Indexicality; Self-Knowledge; Theory Theory; Unconscious

Further Readings

Armstrong, D. (1968). *A materialist theory of the mind.* London, England: Routledge.

Byrne, A. (2005). Introspection. *Philosophical Topics, 33,* 79–104.

Carruthers, P. (2011). *The opacity of mind.* New York, NY: Oxford University Press.

Chalmers, D. (2003). The content and epistemology of phenomenal belief. In Q. Smith & A. Jokic (Eds.), *Consciousness: New philosophical perspectives* (pp. 220–272). New York, NY: Oxford University Press.

Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading.* New York, NY: Oxford University Press.

Gopnik, A. (1993). How we read our own minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16,* 1–14.

Horgan, T., & Kriegel, U. (2007). Phenomenal epistemology: What is consciousness that we may know it so well? *Philosophical Issues, 17/1,* 123–144.

Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge.* Cambridge, MA: Harvard University Press.

Nichols, S., & Stich, S. (2003). *Mindreading: An integrated account of pretense, self-awareness, and understanding of other minds.* New York, NY: Oxford University Press.

Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Ryle, G. (1949). *The concept of mind.* Chicago, IL: University of Chicago Press.

Schwitzgebel, E. (2011). *Perplexities of consciousness.* New York, NY: Oxford University Press.

Smithies, D., & Stoljar, D. (Eds.). (2012). *Introspection and consciousness.* New York, NY: Oxford University Press.

Shoemaker, S. (1996). *The first-person perspective and other essays.* Cambridge, England: Cambridge University Press.

Invisible Hand Explanations

Originally used in the 18th century by Adam Smith in both *An Inquiry Into the Nature and Causes of the Wealth of Nations* (1776) and *The Theory of Moral Sentiments* (1759), the phrase *invisible hand* posits a metaphor that communicates the idea that a beneficial outcome may arise without anyone's intention or design. In this original understanding, to say that some outcome, pattern, or end was brought about by an invisible hand is to say that the result in question carries some benefit (and for that reason appears to have been brought about by design), yet the outcome is not something that the agents intended to bring about. Since the invisible hand refers to the process by which an outcome is generated, it should be possible to set forth *explanations* of such processes: Such "invisible hand explanations" are of particular interest to social science because much of the social realm would seem to be the outcome of unintended action. In the 20th century, Karl Popper, in his attack on all forms of social engineering and historicism, has championed the notion of unintended consequences as a phenomenon considered by him as central to social reality.

Uses

Adam Smith was the first to use the phrase, but he employed it only three times in all of his works. In the *Theory of Moral Sentiments,* Smith suggests that wealthy individuals who seek their own interests will through the employment of labor effectively distribute the same amount of wealth to working individuals as if the land had been equally distributed all along. In the *Wealth of Nations,* Smith contests the practice of restricting imports and contends that in fact individuals prefer domestic over foreign goods, so that the domestic benefit will accrue without the need of restrictions on imports. In these and in other instances, such as the emergence in society of a pattern of specialization (the division of labor), Smith contends that

processes are at work that generate beneficial outcomes without any design, intention, or agreement.

Smith was not the first, or the last, to appeal to such processes. Bernard Mandeville had argued (in *The Fable of the Bees: Or, Private Vices, Public Benefits,* 1714) that prosperity could be achieved by setting in place a rule of law and allowing self-interested individuals to produce, create, and buy and sell. In the 19th century, the Austrian economist Carl Menger explained the genesis of money by invoking an account of how individuals endeavor to trade without the encumbrances of bartering good for good.

In the 20th century, F. A. Hayek (in *The Constitution of Liberty; Law, Legislation, and Liberty*) reiterated the theme of unintended processes, contending that most of the institutions and rules on which a complex civilization depends could only arise via some sort of invisible hand evolution. In political philosophy, Robert Nozick used the idea of the invisible hand to show (in *Anarchy, State, and Utopia*) how a governing state might arise even though no one intended it to emerge. Others have suggested that the emergence of everyday conventions, even those of language (David Lewis, in *Convention*), arise via an unintended or invisible hand process. Another recent use of the notion of an invisible hand guiding a process can be found in theories modeling the advancement of science on economics: for example, Michael Polanyi, explicitly invoking Adam Smith's phrase, suggested, in his article "The Republic of Science: Its Political and Economic Theory" in 1962, that scientific progress can be seen as a spontaneous order resulting from the independent initiatives of scientists.

Explanations

An invisible hand explanation may be called for in any instance in which some relatively complex state of affairs—a rule, pattern, or institution—could have arisen without design or agreement. Although Adam Smith used the original metaphor to suggest that the resultant pattern was beneficial, this is not a necessary part of the explanation. In this sense, an invisible hand explanation is value free: It can be invoked for any complex outcome, regardless of its benefit or harm. The primary kind of invisible hand explanation begins from the description of initial conditions. These would include a description of the circumstances or constraints on action, as well as an account of the agents, their motives and beliefs, as well as their capacity to reason. Given this starting point, the explanation must posit some kind of law or at least a lawlike tendency that would explain why the agents act in the way they do. Finally, the explanation must set forth a narrative in which it is made plain that the outcome was not a part of any agent's intention. The outcome would be shown to emerge via the accumulation or aggregation of individual actions and results. The narrative may reveal that the process occurred over a long period of time or over a narrow slice of time. Many of those who posit such processes suggest that there are evolutionary accounts in which the final outcome occurs over years. One problem with such accounts is that it is crucial to show how the acting agents generated the result in question. Over a long-enough period any outcome could be shown to be an unintended result of human action! Therefore, it is essential to show the causal links between individual actions and the overall outcome.

Eugene Heath

See also Austrian Economics; Complexity and the Social Sciences; Conventions, Logic of; Cooperation, Cultural Evolution of; Cooperation/Coordination; Economics of Scientific Knowledge; Hayek and the "Use of Knowledge in Society"; Individualism, Methodological; Spontaneous Order

Further Readings

Berry, C. (1997). *Social theory of the Scottish Enlightenment.* Edinburgh, Scotland: Edinburgh University Press.

Hamowy, R. (1987). *The Scottish Enlightenment and the theory of spontaneous order.* Carbondale: Southern Illinois University Press.

Hayek, F. A. (1960). *The constitution of liberty.* Chicago, IL: University of Chicago Press.

Hayek, F. A. (1973). *Law, legislation and liberty: Vol. 1. Rules and order.* Chicago, IL: University of Chicago Press.

Hodgson, G. (1994). Hayek, evolution, and spontaneous order. In P. Mirowski (Ed.), *Natural images in economic thought: Markets read in tooth and claw* (pp. 408–447). Cambridge, England: Cambridge University Press.

Lewis, D. (2008). *Convention: A philosophical study.* Cambridge, MA: Harvard University Press. (Original work published 1969)

Mandeville, B. (1988). *The fable of the bees, or private vices, public benefits* (2 vols., with a critical, historical, and explanatory commentary by F. B. Kay). Indianapolis, IA: Liberty Fund. (Original work published 1714)

Nozick, R. (1974). *Anarchy, state, and utopia*. New York, NY: Basic Books.

Nozick, R. (1994). Invisible-hand explanations. *American Economic Review, 84,* 314–318.

Polanyi, M. (1962). The republic of science: Its political and economic theory. *Minerva,* 1, 54–74. Retrieved from http://www.missouriwestern.edu/orgs/polanyi/mp-repsc.htm

Rothschild, E. (2001). The bloody and invisible hand. In *Economic sentiments: Adam Smith, Condorcet, and the Enlightenment* (pp. 116–156). Cambridge, MA: Harvard University Press.

Samuels, W. J., Johnson, M. F., & Perry, W. H. (2011). *Erasing the invisible hand: Essays on an elusive and misused concept in economics.* Cambridge, England: Cambridge University Press.

Ullmann-Margalit, E. (1978). Invisible hand explanations. *Synthese, 39,* 263–291.

Ullmann-Margalit, E. (1997). The invisible hand and the cunning of reason. *Social Research, 64,* 181–198.

Joint Attention and Social Cognition

This entry explains what the capacity for joint attention involves and reviews the main theories about its role in social cognition, as well as theories about its development and mechanism, showing its psychological and philosophical underpinnings.

Joint attention is the capacity enjoyed by humans and, on some accounts, nonhuman primates to attend to objects in their environment together with others. For this to be possible, at least two conditions have to be met: First, the involved creatures have to be attending to the same thing and, second, they have to be aware that they are attending to the same thing. The capacity for joint attention is initially displayed by human infants at the age of around 9 months, when they begin to expand upon earlier face-to-face interactions with their caregivers, through the acts of pointing and diversion of gaze, so as to attract the caregiver's attention to a third object, or through reaction to caregivers' similar acts. It is widely agreed that the display of joint attention marks a key step in infants' cognitive development: It attests to the infant's ability to grasp, and operate within, a triadic epistemic constellation that involves the infant, the caregiver, and the object of attention.

Over the last two decades or so, it has increasingly been recognized that the ability to participate in such constellations is of key relevance for concept acquisition, the ability to distinguish between self and other, the capacity to grasp aspects of another's state of mind, and the ability to share feelings and emotions. It thus occupies a central role in research concerned with social cognition, in a variety of disciplines that include psychology, philosophy of mind, and social neuroscience. The cognitive importance of the phenomenon is highlighted further by the fact that the capacity to jointly attend to objects with others is impaired in persons with autism, which has given rise to medical interest in the topic.

To organize the increasingly rich and varied literature on joint attention, this entry will distinguish between two main issues. The first one is the hotly debated question of just what role joint attention plays in social cognition and how it is acquired. The second question is concerned with the psychological and philosophical frameworks that are meant to explain how joint attention is possible. The entry will address each in turn, and it will conclude with some remarks about the relevance of joint attention for the philosophy of social science.

The Development and Function of Joint Attention

There are two fundamentally distinct views on the place of joint attention in humans' cognitive development. Michael Tomasello has suggested that the capacity for joint attention can be obtained only when the involved persons have acquired the ability to understand each other as intentional agents and to be committed to the joint pursuit of shared goals. This is, on the present view, a uniquely human capacity that gives rise, at around 12 months of

age, to triadic interactions in which the infant plays the leading role. Without the capacity for this kind of commitment, the joint pursuit of goals that are shared between the involved agents is not possible and the knowledge that others choose what to attend to within their visual fields remains unavailable. One immediate consequence of this approach is that only humans can, in principle, be capable of joint attention, since no other creature is thought to possess the range of concepts necessary for meeting the second condition.

The other position, which is held, for instance, by Colwyn Trevarthen, Peter Hobson, Vasu Reddy, and others, holds that meeting the second condition does not depend upon the display of an array of sophisticated concepts. On this view, joint attention plays an important role with regard to the acquisition of a range of social and otherwise cognitive capacities. Thus, one may think that joint attention is necessary (though, according to Ingar Brinck and Peter Gärdenfors, not sufficient) for creatures to engage in collective acts; or it may be supposed that joint attention plays a crucial role in the understanding of spatial perspectives; or one may argue, as Jose Bermudez does, that joint attention gives rise to an understanding of self that plays out in a contrast space with a corresponding understanding of the other creature.

Explanatory Frameworks

The debate about the developmental role of joint attention is closely integrated with, though not identical to, the discussion of which conceptual framework is best suited to explain the phenomenon. What one's views are on the above range of questions will have an impact on which framework one thinks is appropriate in this context, and vice versa. The main contenders here can be grouped into three categories, two of which share a commitment to a representational theory of mind that adherents to the third group reject.

One theory that has been invoked in order to explain joint attention, particularly in connection with issues pertaining to autism research, is the so-called theory theory of mindreading. Simon Baron-Cohen is, as far as joint attention is concerned, perhaps the most prominent defender of this account. He suggests that the capacity for joint attention relies on a set of mental modules that process information in isolation from their environments in a swift and nonconscious way. The capacity to focus on what another creature is focusing on is facilitated by an "eye direction detector" that builds dyadic representations of eye behavior. In humans, this module is coupled with a "shared attention mechanism" that constructs triadic representations, which include an embedded element specifying that other and self are both attending to the same thing. These modules facilitate the development of a theory of mind that allows you to ascribe mental states to yourself and to others.

The second conceptual framework, and probably the most popular one to be invoked in the explanation of joint attention, is Simulation Theory. In the very broadest terms, it explains how persons can come to attend jointly to objects with others in terms of a process of perspective taking. You understand what the other person focuses on by stepping into her shoes, by simulating her point of view. Originally marshalled as a less conceptually demanding alternative to the Theory Theory, the simulationist account has been developed in a variety of ways and with considerable sophistication. Some versions, such as that of Tomasello and his colleagues, are committed to the view that an explicit grasp of the other as intentional agent has to be in place for joint attention to be possible, while other accounts deny that this is necessary. Alvin Goldman distinguishes between "high-level" and "low-level" mindreading; the latter relies on unconscious and conceptually undemanding mental processes. One prominent account is Vittorio Gallese's, who thinks of simulation as an embodied modeling process concerned with the expected motor consequences of an action, which can be mapped onto either the agent or her co-attender. This move is designed to avoid the often raised criticism by defenders of the third conceptual framework that Simulation Theory relies on a conception of self and other that 9-month-old infants, or arguably nonhuman primates, are unlikely to possess.

The third framework really is a collection of views that can be seen as alternatives to both Theory Theory and Simulation Theory and that are united in their rejection of a representational theory of mind. The views contained within this group differ quite significantly in their commitments, however. Shaun Gallagher defends an enactivist view of joint attention, which argues that the meaning of actions that occur in a joint activity is directly present to the

other creature, without reliance on simulative processes of any kind. John Campbell defends a relational view of joint attention, according to which the three-place epistemic relation that obtains between attender, co-attender, and object of attention is to be thought of as primitive. In a joint constellation, the other person (rather than a representation of her) features directly in your experience.

One final approach that could be grouped under the third category but that has been so influential in the debate about joint attention that it may also be seen as an independent set of views comes from developmental psychology and was pioneered, in the 1970s and 1980s, by Colwyn Trevarthen; its most important defenders today are Peter Hobson and Vasu Reddy. The key notion this approach builds on is *intersubjectivity,* the capacity of human infants to attune their states of mind, in social interactions with their caregivers, to others. This sharing of feelings is first evident in what Trevarthen calls "primary intersubjectivity," the face-to-face interactions that develop into episodes of joint attention. Hobson and Reddy both argue, in different ways, that these early encounters amount to a sharing of feelings that carries over into and explains the capacity for joint attention. Joint attention, in this approach, ought to be seen as a fundamentally second-person phenomenon. This line of thought is at odds with both of the two other frameworks outlined previously.

Joint Attention and the Philosophy of Social Science

Joint attention has not so far been recognized as being of key importance for the philosophy of social science. It is quite obviously a cornerstone of human interaction and thus needs to be taken seriously by anyone interested in the psychological foundations of the social domain. However, the social sciences are primarily concerned with the behavior of complex collectives, and phenomena as elusive as joint attention in humans less than one year of age are not governed by the kinds of robust laws that you may expect the social sciences, or at least those that are concerned with the scientific explanation of social phenomena rather than their empathic understanding, to be dealing with. On the other hand, the rapidly increasing amount of work on the foundations of social cognition, in which joint attention plays a crucial role, is now an established research program within which a vigorous conceptual and methodological debate is taking place. The future relevance of joint attention for the philosophy of social science will thus depend, first, on the degree to which the social sciences recognize the importance of social cognition for a comprehensive understanding of the social domain; second, it will also depend on the degree to which they are open to research methodologies that can accommodate insights that aren't always most helpfully expressed in lawlike statements.

Axel Seemann

See also Embodied Cognition; Empathy; Grounded Cognition and Social Interaction; Simulation Theory; Social Cognition; Social Perception; Social Neuroscience; Theory Theory

Further Readings

Bard, K. A., & Leavens, D. A. (2009). Socio-emotional factors in the development of joint attention in human and ape infants. In L. Roska-Hardy & E. M. Neumann-Held (Eds.), *Learning from animals?* (pp. 89–104). Hove, England: Psychology Press.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* Cambridge: MIT Press.

Bermudez, J. L. (1998). *The paradox of self-consciousness.* Cambridge: MIT Press.

Brinck, I., & Gärdenfors, P. (2003). Co-operation and communication in apes and humans. *Mind & Language, 18*(5), 484–501.

Campbell, J. (2002). *Reference and consciousness.* Oxford, England: Oxford University Press.

Campbell, J. (2011). An object-dependent perspective on joint attention. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 415–430). Cambridge: MIT Press.

Eilan, N., Hoerl, C., McCormack, T., & Roessler, J. (Eds.). (2005). *Joint attention: Communication and other minds.* Oxford, England: Oxford University Press.

Gallagher, S. (2005). *How the body shapes the mind.* Oxford, England: Oxford University Press.

Gallagher, S. (2011). Interactive coordination in joint attention. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 293–306). Cambridge: MIT Press.

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences, 4,* 23–48.

Goldman, A. (2005). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford, England: Oxford University Press.

Hobson, P. (2002). *The cradle of thought: Exploring the origins of thinking*. London, England: Macmillan.

Hobson, P. (2005). What puts the jointness into joint attention? In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 185–204). Oxford, England: Oxford University Press.

Hobson, P. (2006). *Foundations for self-awareness: An exploration through autism*. Boston, MA: Blackwell.

Leavens, D. A., Racine, T. P., & Hopkins, W. D. (2009). The ontogeny and phylogeny of non-verbal deixis. In C. Knight & R. Botha (Eds.), *The prehistory of language* (pp. 142–166). Oxford, England: Oxford University Press.

Moll, H., & Meltzoff, A. N. (2011). Joint attention as the fundamental basis of understanding perspectives. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 393–413). Cambridge: MIT Press.

Moore, C., & Dunham, P. J. (1995). *Joint attention: Its origin and role in development*. Hillsdale, NJ: Lawrence Erlbaum.

Reddy, V. (2008). *How infants know minds*. Cambridge, MA: Harvard University Press.

Seemann, A. (Ed.). (2011). *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience*. Cambridge: MIT Press.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28,* 1–61.

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech: The beginning of human communication* (pp. 321–347). Cambridge, England: Cambridge University Press.

Judgment Aggregation and the Discursive Dilemma

This entry explains the important phenomenon of discursive dilemma, which can arise in certain cases of collective judgments in various social institution settings; outlines the main approaches to it; and shows how an interdisciplinary body of research involving philosophy as well as several social sciences has developed around it.

Judgment aggregation theory investigates which procedures a group could or should use to form collective judgments ("yes" or "no") on a given set of propositions or issues, based on the judgments of the group members. How, for instance, should the citizens of a state reach collective judgments on the three propositions—that multiculturalism is desirable, that immigration should be promoted, and that the former implies the latter? And how should a jury in court form collective judgments on the propositions that the defendant has broken the contract, that this contract was legally valid, and that the defendant is liable to pay damages?

The Discursive Dilemma and Political Philosophy

Such collective decision problems are vulnerable to the *discursive dilemma*, a phenomenon generalizing the doctrinal paradox in jurisprudence. The source of the dilemma is that the propositions under consideration are logically interconnected. In our first example, the third proposition is a conditional involving the first two propositions; and in our second example, the third proposition is equivalent to the conjunction of the first two propositions (according to the generally acknowledged legal doctrine that breach of a valid contract is necessary and sufficient for liability).

The problem is that the initially most natural and democratically appealing procedure—proposition-wise majority voting—may generate inconsistent collective judgments. In the case of our first example, Table 1 illustrates a situation in which the population is split into three camps such that, overall, a majority believes that multiculturalism is desirable (proposition *P*), another majority believes that *if* multiculturalism is desirable *then* immigration should be promoted (proposition *if-P-then-Q*), and yet another majority believes that immigration should *not* be promoted (*not-Q*).

Thus, this phenomenon of majoritarian inconsistencies poses a serious challenge to the very meaning and possibility of democracy, since it seems that collective judgments cannot be both consistent and democratically responsive to people's judgments.

To restore collective rationality, two routes are often contrasted. Under the *premise-based* route, the collective adopts the majority-supported judgments *P* and *if-P-then-Q* (interpreted as two *premises*), from

Table 1 Inconsistent Majority Judgments

	P	*If-P-then-Q*	*Q*
1/3 of the population	Yes	Yes	Yes
1/3 of the population	Yes	No	No
1/3 of the population	No	Yes	No
The majority	Yes	Yes	No

Source: Author.

which it derives the judgment *Q* (interpreted as a *conclusion*). Under the *conclusion-based* route, the collective instead adopts the majority-supported judgment *not-Q,* and it either forms no judgments at all on the premise propositions or forms some judgments on them that are logically consistent with *not-Q,* such as the judgments *not-P* and *if-P-then-Q*. In short, the premise-based approach respects majorities on premises while overruling majorities on conclusions, while the conclusion-based approach does the converse.

In response to the discursive dilemma, a highly interdisciplinary body of research has developed, conducted mainly by economists, philosophers, political scientists, and computer scientists. The less formal branch of research has its home in political philosophy. It focuses on the nature and role of the collective agent and the extent to which such an agent should provide reasons (premises) for its policies (conclusions). As is sometimes argued from the perspective of republican democratic theory, the state must act upon and publicly provide reasons for its actions to be contestable. Contestability of state actions is in turn important for preventing arbitrary state interference in citizens' lives, that is, to render citizens free in the republican sense. The need for reason-based state actions or policies is often taken to imply the superiority of premise-based over conclusion-based aggregation.

The Formal Theory of Judgment Aggregation

The more formal area of research stands in the tradition of Arrovian Social Choice Theory. The judgment aggregation problem is formulated in full abstract generality. Two central ingredients of the theory are, first, the group's *agenda,* that is, the set of propositions on which judgments are formed, and second, the notion of an *aggregation rule* or *procedure,* that is, a function that takes each person's set of judgments as input and returns a collective set of judgments. Simple examples of aggregation rules are proposition-wise majority rule, proposition-wise quota rules (with acceptance thresholds that may differ from the majority threshold and may vary across propositions), premise-based voting, conclusion-based voting, and the "expert rule," which universally adopts the judgments of a fixed individual (the "expert" or "dictator"). The generality of the framework stems from the fact that virtually any kind of decision can be construed as the formation of judgments on particular propositions. Notably, the classical preference aggregation problem in Social Choice Theory emerges as a special case, because a preference relation can be construed as a set of judgments on propositions of the form "*x* is better than *y,*" where *x* and *y* denote choice alternatives.

The Axiomatic Approach

Within judgment aggregation theory, one may broadly distinguish between an *axiomatic* and a *constructive* approach, the two of which go hand in hand. The axiomatic approach starts by formulating general requirements ("axioms") on aggregation rules that capture normative principles or intuitions. An example of an axiom of procedural fairness is *anonymity,* which forbids differential treatment of voters—in other words requires that the collective judgment set only depends on the number of individuals holding each given judgment set, regardless of their identity. This axiom, for instance, excludes the aforementioned expert rule. The axiom of consistency of collective judgment sets excludes proposition-wise majority rule, as the discursive dilemma shows. Once a set of axioms is specified, one proceeds by determining all the judgment aggregation rules satisfying the axioms, a more or less difficult mathematical exercise. Ideally, there is a single such rule, but often there are *many* rules (leaving a choice to be made) or *no* rules (forcing one to abandon an axiom). Indeed, in a series of *impossibility theorems,* it has been established that various combinations of axioms are not satisfied by any aggregation rule if the agenda of propositions is sufficiently complex. Many of these theorems are in a similar spirit to Kenneth Arrow's famous impossibility theorem in preference aggregation theory. Indeed, one of them stands out as being an exact generalization of Arrow's theorem from preference aggregation problems to arbitrary-judgment aggregation problems.

A quite different impossibility theorem generalizes Amartya Sen's influential "Impossibility of a Paretian Liberal"; it brings to light a conflict between respecting unanimous judgments and respecting the right of individuals or minorities to alone determine the collective judgment on propositions within the private sphere or the area of special competence.

The Constructive Approach

This approach tries directly to devise concrete aggregation rules for reaching consistent and democratically responsive collective judgments, without a preceding axiomatic derivation. The following salient proposals or paradigms can be contrasted: (a) *premise-* or *conclusion-based* aggregation rules; (b) *quota rules* with well-calibrated acceptance thresholds; (c) *sequential* rules, where the propositions in the agenda are voted upon one by one in an order of priority and where the vote on any proposition is suspended if the previous voting outcomes on propositions of higher priority already imply a judgment ("yes" or "no") on the current propositions; (d) *distance-based* rules, where the collective adopts a consistent set of judgments whose sum-total distance to people's sets of judgments is as small as possible, with respect to some distance measure between judgment sets; and (e) *scoring* rules, where the collective adopts a consistent set of judgments that receives maximal sum-total score from the individuals, with respect to some definition of "scores."

Localistic Versus Holistic Aggregation

Under a *localistic* (or *proposition-wise*) understanding of democratic responsiveness, the collective judgment on any given proposition should be formed solely on the basis of people's judgments on this proposition, independently of their judgments on other propositions. By contrast, under a *holistic* conception of democratic responsiveness, the collective judgment on, say, whether immigration should be promoted may be influenced by people's judgments on other propositions, such as "premises" or even unrelated propositions about taxation. Here, even an overwhelming majority judgment on whether immigration should be promoted may be overruled in the name of people's judgments elsewhere.

Localism is the content of the (controversial) *independence* axiom, the counterpart in judgment aggregation theory of the (equally controversial) axiom of "independence of irrelevant alternatives" in preference aggregation theory. A virtue of independence is that it is necessary for preventing the manipulation of outcomes through certain types of strategic voting or strategic agenda setting, as was proved. However, independence features as the central axiom in most impossibility theorems. Hence, the goal of aggregating localistically is unachievable (for agendas subject to the impossibility result)—whether or not localism is normatively desirable.

The Procedural Versus the Epistemic Approach

Two contrasting approaches or aims may be pursued when designing the aggregation rule. The *procedural* approach aims for a "procedurally fair" rule; for instance, anonymity is usually a central procedural virtue. The *epistemic* approach aims for a rule that generates "correct" or "true" collective judgments by an external, procedure-independent standard of correctness or truth; here, anonymity may be violated in the name of differences in information or competence. While most of the literature has a proceduralist flavor, some work takes the epistemic perspective and stands in the tradition of Condorcet's jury theorem.

Franz Dietrich

See also Collective Agents; Collective Rationality; Decision Theory; Group Beliefs; Sen's Paretian Liberal; Social Choice Theory; Social Epistemology

Further Readings

Dietrich, F. (2007). A generalised model of judgment aggregation. *Social Choice and Welfare, 28,* 529–565.

Dietrich, F., & List, C. (2007). Arrow's theorem in judgment aggregation. *Social Choice and Welfare, 29,* 19–33.

List, C., & Pettit, P. (2002). Aggregating sets of judgments: An impossibility result. *Economics and Philosophy, 18,* 89–110.

List, C., & Pettit, P. (2011). *Group agency: The possibility, design and status of corporate agents.* New York, NY: Oxford University Press.

Nehring, K., & Puppe, C. (2010). Abstract Arrowian aggregation. *Journal of Economic Theory, 145,* 467–494.

Pettit, P. (2001). Deliberative democracy and the discursive dilemma. *Noûs, 35*(11), 268–299.

Kinds: Natural Kinds Versus Human Kinds

Both the natural and the social sciences posit taxonomies or classification schemes that divide their objects of study into various categories. Many philosophers hold that what makes some taxonomic schemes more legitimate than others is that they correspond to actually existing divisions in nature, which they label "natural kinds." In other words, some classification schemes "carve nature at the joints" (to use a phrase inspired by Plato), while others are merely arbitrary or gerrymandered. In the natural world, examples of natural kinds would be *insect* and *metal,* but not *bug* and *rock,* or *insects-born-on-Monday* and *metals-whose-names-start-with-"s,"* which are nonnatural or "artificial" kinds.

When it comes to the social sciences, the question is "Do some categories correspond to social kinds or human kinds (the two terms are often used interchangeably), just as some categories in the natural sciences correspond to natural kinds?" Social scientists posit kinds of human being (e.g., *consumer, psychopath*), kinds of social institution (e.g., *political party, economic market*), kinds of social process (e.g., *ritual, immigration*), and so on. But is it enough for a social scientist simply to invoke a human or social category to conclude that such a kind really exists and that it ought to be admitted into our ontology?

This entry begins by examining purported differences between natural kinds and human or social kinds. Then, whether or not human kinds are fundamentally different from natural kinds, it will ask how we can distinguish genuine human kinds from spurious ones. Finally, the issue of natural kinds and human kinds will be related to the common claim that certain categories in both the natural and the social sciences are "social constructs."

Differences Between Natural Kinds and Human Kinds

Recent philosophical work has tended to favor an essentialist understanding of what it is for a category to correspond to a natural kind. Although essentialists are not united on the precise features that characterize natural kinds, they tend to posit that each natural kind is characterized by all or some of the following: (a) properties that are necessary and sufficient for membership in the kind, (b) microstructural properties, (c) intrinsic properties, (d) modally necessary properties, and (e) properties that are discoverable by science. If one understands natural kinds in this essentialist fashion, then it is fairly clear that most if not all social categories will fail to qualify as natural kinds. Without going into these features in detail, it would appear that many social categories are not definable in terms of necessary and sufficient properties, are not characterized by an underlying microstructure, are not wholly determined by their intrinsic properties, and do not have their properties as a matter of metaphysical necessity. Moreover, if they are discoverable by science, it will be a social rather than a natural science. Hence, this essentialist account of natural kinds would seem to be a nonstarter for the social domain.

But the essentialist account of natural kinds, though currently popular among some philosophers, is not universally held. Without taking the social sciences into account, the essentialist consensus is showing signs of unraveling under the pressure of categories in biology and many of the other "special sciences" (e.g., geology, biochemistry, etc.). Still, even if one does not endorse an essentialist understanding of natural kinds, there may be fundamental differences between natural kinds and human or social kinds.

The first obstacle to assimilating human kinds to natural kinds lies in the fact that real entities are often distinguished from artificial ones on the grounds that the latter are mind dependent or the result of human artifice. But if we adopt that criterion, then we would be led to judge all human or social kinds to be nonreal. Even though they may not be intentionally and consciously produced by human beings, all human or social kinds are influenced in some way by human action and social forces and cannot be regarded as independent of human beings and their minds. This seems to be one obvious way in which human kinds are different from natural kinds.

A second difference between human and natural kinds has been elaborated by Ian Hacking, who argues that human kinds are "interactive" or subject to the "looping effect." Hacking has illustrated this claim with various categories, such as *child abuse, multiple-personality disorder, fugue* ("mad traveler" syndrome), and others. Consider the category of *child abuse*. Once this practice is identified and labeled, those who engage in it may alter their behavior either by refraining from the practice altogether or by engaging in it more covertly or in some other way. In these cases, the phenomenon may be altered as a result of human intervention, for instance, by becoming less visible. Since the very nature of the phenomenon has changed, it then "loops back" to influence our beliefs about it, which may in turn have further effects on this practice, and so on. Though looping effects undoubtedly influence the nature of social phenomena, it is not clear that such effects are confined to the social realm. For instance, some have suggested that biological species that are a product of artificial selection can be subject to looping effects too, such as the domestic dog (*Canis familiaris*), whose phenotypical traits have been shaped by the beliefs and actions of human beings.

A third difference between natural kinds and human or social kinds has been emphasized by John Searle, who says that what it is for some social kind *x* to be *x* is simply to be regarded as *x*. For example, what it is for something to be money is for it to be regarded as money, to be used as money, and to be believed to be money. Hence, social kinds are ontologically subjective, being dependent for their very existence on human attitudes toward them, which distinguishes them from natural kinds. But critics have observed that Searle's account seems to apply only to the most conventional of social kinds, such as *money* or *government,* which depend for their very existence on human attitudes concerning those very categories. Other social kinds, such as *racism* or *economic recession,* do not seem to depend on our having attitudes toward them at all. They are therefore arguably not ontologically subjective, at least not in the same sense as kinds such as *money.*

Finally, perhaps the most widely cited difference between natural kinds and human kinds pertains to their purported value-ladenness or their alleged normative dimension. Some social scientists and philosophers regard many social categories to be evaluative and consider that their range of application and the properties associated with them are not determined solely by the nature of the social phenomena themselves but at least partly by our moral and ideological attitudes toward those social phenomena. If categories like *race, gender,* and *child abuse* are shaped by our value judgments or ideological stances, then that may set them apart from categories denoting natural kinds. But even those social theorists who insist that the very aims of social science ought to be normative (e.g., aiming at human empowerment or emancipation) seem to distinguish between the descriptive and normative dimensions of social theorizing. If so, this leaves room for the possibility that at least some of our social categories can serve a more purely descriptive and explanatory purpose.

Genuine and Spurious Human Kinds

Whether or not there are fundamental differences between natural kinds and human or social kinds, it is possible to maintain that at least some social categories correspond to social or human kinds. The next challenge consists in saying just which categories delineate genuine as opposed to spurious human

kinds, or how to distinguish between real and nonreal kinds in the social domain.

One promising route might be to return to one of the original sources of the notion of a natural kind (or "real kind") in the work of the 19th-century English philosopher John Stuart Mill. Mill was skeptical as to whether races constituted real kinds of humans that would constitute species or subspecies of human beings. His reasoning depended on the idea that a real or natural kind, whether in the natural or social sciences, ought to be characterized by a multitude of properties that are not simply deducible from one another. These properties ought to be scientifically important in the sense of being explanatory and playing a role in inductive inference. Moreover, he accorded the primary responsibility for determining whether this was the case to the investigators in each branch of knowledge, also accepting that there may be different classification schemes deployed by theorists in different scientific disciplines or subdisciplines. If we follow Mill broadly in this approach, we may consider human kinds to correspond to those categories identified by social scientists and investigators studying the properties of human beings and human society. The categories that they arrive at as a result of their investigations, provided they are genuinely explanatory and feature in inductive inference, would be the ones that correspond to real human kinds, while those that are not can be safely dismissed. Also following Mill, there need be no unique classification of human beings, their institutions, processes, and so on, into a hierarchy of categories. Rather, we may end up classifying humans into many crosscutting systems of kinds (e.g., in terms of both *class* and *ethnicity*) without undermining the reality of those kinds.

Social Kinds and Social Construction

What should we make of ubiquitous claims that at least some social kinds are "social constructions"? There is a sense in which any such claim about a social kind is vacuous, since social processes, institutions, and attitudes are all in some way constructed by human society. But is there a more interesting sense in which such claims are warranted? There would seem to be at least two possibilities. The first is that a socially constructed kind is one that is more deliberately a result of human artifice than might appear at first. (This may occur along the lines of Hacking's looping effect or in the manner of Searle's conventional social kinds.) In connection with this, social constructionists may provide informative accounts of the way in which the process of construction took place, tracing it back to particular historical eras or identifying it with certain social movements with specific ideologies or interests. A second possibility is that saying that a kind is a social construct is opposed to regarding it as biological or physiological in nature. Such claims are often made with regard to kinds that are commonly thought to be in the domain of the natural sciences but should rather be regarded as pertaining to the social sciences, such as *race* and *gender*. In such cases, to claim that a kind is a social construct is to say that the basis of the kind is not biological but can be found in social processes and relations instead.

Conclusion

Categories in the social sciences may be more mind dependent, interactive, conventional, and normative than those in the natural sciences, but that may not prevent them from corresponding to genuine human or social kinds. Moreover, genuine human kinds may be distinguished from spurious kinds on the basis of features such as their role in inductive inference and their explanatory value. This may enable us to better address some questions regarding categories that arise in the practice of social science, such as the reality of the category of *race* or how many *gender* categories there are in the human species.

Muhammad Ali Khalidi

See also Essentialism; Metaphysics and Science; Race, Theories of; Relativisms and Their Ontologies; Searle and the Construction of Social Reality; Social Construction of Reality; Social Constructivism; Social Ontology, Recent Theories of

Further Readings

Griffiths, P. E. (2004). Emotions as natural and normative kinds. *Philosophy of Science, 71,* 901–911.

Hacking, I. (1999). *The social construction of what?* Cambridge, MA: Harvard University Press.

Haslanger, S. (2006). What good are our intuitions? Philosophical analysis and social kinds. *Aristotelian Society Supplementary Volume, 80,* 89–118.

Khalidi, M. A. (2010). Interactive kinds. *British Journal for the Philosophy of Science, 61,* 335–360.
Mallon, R. (2007). Human categories beyond non-essentialism. *Journal of Political Philosophy, 15,* 146–168.
Mill, J. S. (1974). *A system of logic ratiocinative and inductive.* Toronto, Ontario, Canada: University of Toronto Press. (Original work published 1843)
Millikan, R. G. (1999). Historical kinds and the "special sciences." *Philosophical Studies, 95,* 45–65.
Searle, J. (1995). *The construction of social reality.* New York, NY: Free Press.
Thomasson, A. (2003). Realism and human kinds. *Philosophy and Phenomenological Research, 67,* 580–609.

Knowing-How Versus Knowing-That

This entry introduces the epistemic distinction between *knowing-how* and *knowing-that* and its philosophical provenance and goes on to present the various positions and nuances regarding that distinction in recent analytical philosophy.

The British philosopher Gilbert Ryle famously highlighted the distinction between knowing that something is the case (e.g., knowing *that* Antarctica is a continent) and knowing how to do things (e.g., knowing *how* to traverse Antarctica). This distinction is sometimes said to be related to others, for example, *epistêmê* and *technê,* reflection and habit/practice/craft, explicit knowledge and tacit/implicit knowledge, and declarative knowledge and procedural knowledge. However, the precise relations between these distinctions is contentious and a matter of substantive debate.

There are at least two ways in which knowing-that and knowing-how might be considered distinct. First, they might be said to be *inequivalent* (perhaps because not all instances of knowing-that are instances of knowing-how). Second, and more strongly, they might be said to be *exclusive* or *strongly contrastable.* The second is stronger in that it entails, but is not entailed by, the first.

The stronger claim was the centerpiece in Ryle's attack on "intellectualism," a position that views internal, nonovert intellectual states (e.g., belief, opinion, knowledge-that, and other propositional or factual attitudes) and intellectual operations (e.g., reasoning) as the key to intelligence and intelligent action, understood to include skill and expertise. Ryle contended that intellectualism entails a "vicious regress," which he proposed to avoid by maintaining that knowing-how, unlike knowing-that, is a type of power to act (e.g., an ability or disposition to behavior), hence his anti-intellectualism, which views such a power to act—rather than propositional attitudes or reasoning—as the key to intelligence and intelligent action.

It is, however, important to keep separate the following four Rylean theses (the first two we have already distinguished):

1. Knowing-how and knowing-that are not equivalent.
2. Knowing-how and knowing-that are strongly contrastable.
3. Knowing-how is a power to act.
4. Knowing-how is the key to intelligence and intelligent action.

In Ryle's view, (4) is a neglected but important truth, to be explained in terms of (3); the latter can be seen as a further specification of (2), which as noted above entails (1). But these four theses can come apart. For example, one might reduce all knowing-that to (a type of) knowing-how and all knowing-how to (a type of) ability: One version of such a position would accept (3) and (4) while rejecting (1) and (2).

Intellectualism, by contrast, rejects both (2) and (3). In defense of (2), it is argued that knowing-how lacks the "standard marks" of knowing-that, such as expressibility, explicitness, justification, and nonaccidental truth. In defense of (3), it is argued that intellectualist alternatives "overintellectualize" everyday knowing-how, skill, and expertise, as well as the mental lives of infants and nonhuman animals. And there is also Ryle's regress argument, mentioned above.

Yet while (2) and (3) were long regarded as orthodoxy, it is fair to say that at the time of writing, the increasingly dominant (though by no means consensus) view among epistemologists working on knowing-how is that these theses cannot be sustained.

Several reasons have been offered for this intellectualist conclusion. First, Paul Snowdon argues that many instances of knowledge-how involve

substantive knowledge-that; for example, knowing *how* to get from London to Swansea before midday depends on knowing *that* one first catches the 7.30 a.m. train to Reading from Paddington, then one . . . , etc." Second, a standard approach to the syntax and semantics of embedded questions within contemporary linguistics treats ascriptions of knowing-how as equivalent to ascriptions of knowing-that: Thus, in an influential paper published in 2001, Jason Stanley and Timothy Williamson argue that "S knows how to A" is true if and only if, for some W such that W is a way of A-ing, S knows *that* W is a way to A. Third, several counterexamples have been offered against (3). To illustrate, King describes a ski instructor who knows how to perform ski stunts, which he teaches his students, even though he is not able to do them himself, and Katherine Hawley describes a hiker who does not know how to escape avalanches although, given her "accidental success" when an avalanche occurs, she is nevertheless able to do so.

Other considerations motivating rejection of (2) and (3) include the causal-explanatory significance of rules and symbolic representations in cognitive science (see Fodor, 1968), the differences between practical knowledge and mere "knacks," and the role of know-how in (performing, learning, practicing) intentional action and our knowledge thereof.

Such intellectualist challenges are compatible with (1). How do they cohere with the datum, which seems to underlie (4), that knowing-how is intimately tied to action? It is open to intellectualists to treat some instances of knowing-that—and intellectual states or operations more generally—as likewise intimately tied to action. From this perspective, (2) and (3) fail not because knowing-how is "theoretical," or action-neutral, but rather because intellectual states such as knowing-that are sometimes "practical," or action oriented.

But we must keep separate the position that (2) and (3) are false from the following, further thesis:

5. Knowing-how is (a type of) knowing-that.

It is often assumed that there are only two options: Knowledge-how is a power to act (3), or it is knowledge-that (5). However, there may be other possibilities, for example, knowing-how might be *familiarity with a practical universal* or a kind of *nonpropositional understanding*. Such views reject (2) and (3) while also denying (5).

Debate over knowing-how and knowing-that has influenced philosophical discussion of, for example, linguistic, logical, moral, and experiential knowledge and has the potential to inform research in psychology and cognitive science (e.g., skills learning, artificial intelligence), linguistics, and theory of education, among other areas.

John Bengson

See also Action, Philosophical Theory of; Agency; Behaviorism, Philosophical Conception of; Causes Versus Reasons in Action Explanation; Epistemology; Philosophy of Expertise; Tacit Knowledge

Further Readings

Bengson, J., & Moffett, M. (2007). Know-how and concept possession. *Philosophical Studies, 136,* 31–57.

Bengson, J., & Moffett, M. (2011a). Nonpropositional intellectualism. In J. Bengson & M. Moffett (Eds.), *Knowing how: Essays on knowledge, mind, and action* (pp. 161–195). Oxford, England: Oxford University Press.

Bengson, J., & Moffett, M. (2011b). Two conceptions of mind and action: Knowledge how and the philosophical theory of intelligence. In J. Bengson & M. Moffett (Eds.), *Knowing how: Essays on knowledge, mind, and action* (pp. 3–58). Oxford, England: Oxford University Press.

Bengson, J., Moffett, M., & Wright, J. (2009). The folk on knowing how. *Philosophical Studies, 142,* 387–401.

Brogaard, B. (2011). Knowing how: A unified approach. In J. Bengson & M. Moffett (Eds.), *Knowing how: Essays on knowledge, mind, and action* (pp. 136–160). Oxford, England: Oxford University Press.

Cath, Y. (2011). Knowing how without knowing that. In J. Bengson & M. Moffett (Eds.), *Knowing how: Essays on knowledge, mind, and action* (pp. 113–135). Oxford, England: Oxford University Press.

Devitt, M. (2011). Methodology and the nature of knowing how. *Journal of Philosophy, 108,* 205–218.

Fodor, J. (1968). The appeal to tacit knowledge in psychological explanation. *Journal of Philosophy, 65,* 627–640.

Ginet, C. (1975). *Knowledge, perception, and memory.* Dordrecht, Netherlands: Reidel.

Hawley, K. (2003). Success and knowledge how. *American Philosophical Quarterly, 40,* 19–31.

Hetherington, S. (2006). How to know (that knowledge-that is knowledge-how). In S. Hetherington (Ed.), *Epistemology futures* (pp. 71–94). Oxford, England: Oxford University Press.

Hornsby, J. (2011). Ryle's *knowing-how,* and knowing how to act. In J. Bengson & M. Moffett (Eds.), *Knowing how: Essays on knowledge, mind, and action* (pp. 80–100). Oxford, England: Oxford University Press.

Kumar, V. (2011). In support of anti-intellectualism. *Philosophical Studies, 152,* 135–154.

Noë, A. (2005). Against intellectualism. *Analysis, 65,* 278–290.

Ryle, G. (1945). Knowing how and knowing that. *Proceedings of the Aristotelian Society, 46,* 1–16.

Ryle, G. (1949). *The concept of mind.* Chicago, IL: University of Chicago Press.

Snowdon, P. (2004). Knowing how and knowing that: A distinction reconsidered. *Proceedings of the Aristotelian Society, 104,* 1–29.

Snowdon, P. (2011). Rylean arguments: Ancient and modern. In J. Bengson & M. Moffett (Eds.), *Knowing how: Essays on knowledge, mind, and action* (pp. 59–79). Oxford, England: Oxford University Press.

Stanley, J. (2011). *Know how.* Oxford, England: Oxford University Press.

Stanley, J., & Williamson, T. (2001). Knowing how. *Journal of Philosophy, 98,* 411–444.

Williams, J. N. (2008). Propositional knowledge and know-how. *Synthese, 164,* 133–155.

Knowledge Society

This entry introduces the emergence of novel forms of postindustrial society, known as *knowledge society,* explains the new place of knowledge in social life as well as the rise of knowledge economies and reviews several repercussions for social and political conduct in such new forms of societies.

Background

The foundations for the transformation of modern societies into "knowledge societies" continues to be based, as was the case for *industrial society,* on changes in the structure of the economies of advanced societies. Economic capital—or, more precisely, the source of economic growth and value-adding activities—increasingly relies on *knowledge.* The transformation of the structures of the modern economy by knowledge as a productive force constitutes the "material" basis and justification for designating advanced modern society as a *knowledge society.*

The significance of knowledge grows in all spheres of life and in all social institutions of modern society. The historical emergence of knowledge societies represents not a revolutionary development but rather a gradual process during which the defining characteristics of society change and new traits emerge. Until recently, modern society was conceived primarily in terms of property and labor. While the traditional attributes of labor and property certainly have not disappeared entirely, a *new* principle, *knowledge,* has been added, which, to an extent, challenges as well as transforms property and labor as the constitutive mechanisms of society.

Knowledge About Knowledge

Knowledge may be defined as a *capacity for action.* The definition of knowledge as capacity for action indicates that implementation of knowledge is open and is dependent on or embedded within the context of specific social, economic, and intellectual conditions.

Knowledge is a peculiar entity with properties unlike those of commodities or of secrets, for example. Knowledge exists in *objectified* and *embodied* forms. If sold, it enters other domains—and yet it remains within the domain of its producer. Unlike money, property rights, and symbolic attributes such as titles, knowledge cannot be transmitted instantaneously. Its acquisition takes time and often is based on intermediary cognitive capacities and skills. Despite its reputation, knowledge is virtually never uncontested. *Scientific* and *technical* knowledge is uniquely important in modern social systems because it produces *incremental* capacities for social and economic action that may be "privately appropriated," at least temporarily. Knowledge has, of course, always had a major function in social life. Social groups, social situations, social interaction, and social roles all depend on, and are mediated by, knowledge. *Power* too has frequently been based on knowledge advantages, not merely on physical strength.

The Knowledge-Based Economy

The emergence of knowledge societies signals, first and foremost, a radical transformation in the *structure of the economy.* What changes are the dynamics of the supply and demand for primary products or raw materials, the dependence of employment on production, the importance of the manufacturing sector that processes primary products, the role of manual labor and the social organization of work, the role of international trade in manufactured

goods and services, the function of time and place in production, and the nature of the limits to economic growth. The common denominator of the changing economic structure is a shift away from an economy driven and governed by *material* inputs into the productive process and its organization toward an economy in which the transformations of productive and distributive processes are increasingly determined by symbolic or knowledge-based inputs.

Social and Political Conduct in Knowledge Societies

The transformation of modern societies into knowledge societies has profound consequences apart from those of its economic system. One of the more remarkable consequences is the extent to which modern societies become *fragile* societies. The fragility of modern societies is a unique condition. Modern societies tend to be fragile from the viewpoint of those large and once-dominant social institutions (e.g., the state, the economy, science) that find it increasingly difficult to impose their will on all of society. Societies are fragile because individuals and small groups are capable, within certain established rules, of asserting their own interests by opposing or resisting the—not too long ago—almost unassailable monopoly of truth by major societal institutions. That is to say, legitimate cultural practices based on the enlargement and diffusion of knowledge enable a much larger segment of society to effectively oppose power configurations that turned out or are apprehended to be tenuous and brittle.

Knowledge societies are (to adopt a phrase from Adam Ferguson) the result of human action, but often not of deliberate human design. Knowledge societies emerge as adaptations to persistent but evolving needs and the changing circumstances of human conduct.

Modern societies are also increasingly *vulnerable* entities. More specifically, the economy as well as the communication or traffic systems are vulnerable to malfunctions of self-imposed practices typically designed to avoid breakdowns. Modern infrastructures and technological regimes are subject to accidents, including large-scale disasters, as the result of fortuitous, unanticipated human action; to nonmarginal or extreme natural events that may dramatically undermine the taken-for-granted routines of everyday life in modern societies; or to deliberate sabotage.

Present-day social systems may be seen to be fragile and vulnerable entities in yet another sense. Such fragility results from conduct as well as the deployment of artifacts designed to stabilize, routinize, and delimit social action. An example is the so-called computer trap or, more generally, the unintended outcomes of intentional social action. In the process of even more deeply embedding computers into the social fabric of society, that is, redesigning and reengineering large-scale social and socio-technical systems in order to manage the complexities of modern society, novel risks and vulnerabilities are created.

Among the major but widely invisible social innovations in modern society is the immense growth of the "civil society" sector. This sector provides an organized basis through which citizens can exercise individual initiative in the private pursuit of public purposes. One is therefore able to interpret the considerable enlargement of the informal economy and also corruption and the growth of wealth in modern society, as well as the increasing but often unsuccessful efforts to police these spheres, as evidence of the diverse as well as expanded capacity of individuals, households, and small groups to take advantage of and benefit from contexts in which the degree of social control exercised by larger (legitimate) social institutions has diminished considerably.

The future of modern society no longer mimics the past to the extent to which this has been the case. History will increasingly be full of unanticipated incertitudes, peculiar reversals, and proliferating surprises, and we will have to cope with the ever greater speed of significantly compressed events. The changing agendas of social, political, and economic life as the result of our growing capacity to make history will also place inordinate demands on our mental capacities and social resources.

Nico Stehr

See also Evidence-Based Policy; Information Society; Philosophy of Expertise; Postindustrial Society; Power; Risk; Technoscience and Society

Further Readings

Bell, D. (1973). *The coming of post-industrial society: A venture in social forecasting*. New York, NY: Basic.

Drucker, P. (1968). *The age of discontinuity: Guidelines to our changing society*. New York, NY: Harper & Row.

Stehr, N. (1994). *Knowledge societies*. Thousand Oaks, CA: Sage.

Stehr, N. (2001). *The fragility of modern societies: Knowledge and risk in the Information Age*. Thousand Oaks, CA: Sage.

Kuhn and Social Science

This entry presents the influences the social sciences had on Thomas Kuhn's theorizing about science, Kuhn's view on the relationship between the natural and social sciences, his influence on the social sciences generally, and, finally, his influence on the sociology of science.

How the Social Sciences Influenced Kuhn

In *The Structure of Scientific Revolutions,* Thomas Kuhn develops a general theory of scientific change. According to Kuhn, scientists working in a field can only make significant progress after they settle on a *paradigm,* a concrete scientific achievement that provides a template for future research in the field. Johannes Kepler's mathematical model for the orbit of Mars is a typical paradigm. It solved an outstanding research problem, providing a model for determining the location of Mars. Furthermore, Kepler's solution provided a general schema for modeling related phenomena, including the orbits of other planets and the orbit of the moon. As far as Kuhn is concerned, a scientific field only becomes mature when a paradigm becomes widely accepted. Only then can the field make the sort of progress we associate with the growth of science.

Kuhn made this insight about the role of paradigms in science while he was working at the Center for Advanced Studies in the Behavioral Sciences. Interacting with social scientists, he discovered the important role paradigms play in research in the natural sciences. Paradigms, Kuhn claimed, are largely absent in the social sciences. Kuhn thus believes that the social and natural sciences are fundamentally different.

Kuhn's Philosophy of the Social Sciences

Though Kuhn never claimed to have any special expertise on the social sciences, he reflected on the relationship between the social and natural sciences later in his career. He believed that the traditional view, a view associated with the sociologist Max Weber, was mistaken. According to this view, the key difference between the natural sciences and the social sciences is that the latter are interpretative whereas the former are not. The former "explain" observable natural facts, while the latter "understand/interpret" meaningful human action. As a result, the objects studied by the social sciences are subject to change as we theorize about them. The objects studied by natural scientists, like cells and atoms, are indifferent to our theorizing.

Kuhn believed that this alleged difference between the natural and social sciences had been exaggerated. The objects studied by natural scientists are not as ready-made as this popular view implies. The natural world is not merely awaiting our discovery. Rather, Kuhn believed that the objects studied by natural scientists are partly constituted by their theorizing. Different theories lead us to understand objects in the natural world differently. Observation is thus theory-laden. Different theories of the atom, for example, will lead us to "see" the world in different ways and to work in different worlds. Kuhn is not claiming there are no significant differences between the natural and social sciences. Rather, his point is that Weber's approach, and that of the British philosopher Peter Winch, to understanding rule-following behavior as fundamentally opposite to natural-scientific inquiry have misidentified what the key difference is.

Kuhn's Influence on the Social Sciences

Kuhn's account of scientific change, outlined above, was greeted with mixed reactions. Many philosophers of science, concerned only with the natural sciences, regarded Kuhn's theory of scientific change as a threat to the rationality of science. Kuhn seemed to suggest that the change from one paradigm to another in a field cannot be made in a rational manner for there are no paradigm-transcending standards by which to evaluate competing paradigms. This was widely regarded as a serious threat to the epistemic integrity of science.

Social scientists responded quite differently to Kuhn's theory of scientific change. Many thought that Kuhn was suggesting that no field was fully scientific unless it had a paradigm. This led many reflective social scientists to search for the paradigms in their fields. This search created some disappointments as social scientists in a number of fields discovered that they did not in fact have a paradigm. Some believed that this offered insight into why

the social sciences do not progress like the natural sciences. Others took it as an indication that their fields were in a preparadigm stage, still awaiting the creation of their first paradigms. In time, as Kuhn's general account of scientific change was challenged, concern about the scientific status of the social sciences passed.

Kuhn and the Sociology of Science

The social-scientific field most profoundly influenced by Kuhn is the sociology of science. Kuhn's *Structure* led to the development of the Strong Program in the Sociology of Scientific Knowledge, with Barry Barnes and David Bloor as its protagonists. The strong program was built on a particular reading of Kuhn's theory of concept application, a view they called "finitism." According to the strong program, every act of classifying an object is underdetermined by evidence and logic. This makes room for the influence of social factors on science. The strong program had a profound influence on developments in the sociology of science, leading to important investigations into how various social factors influence the content of science.

Kuhn was quite uncomfortable with the way sociologists of science interpreted his theory. He distinguished his own view from finitism, insisting that nature plays a far greater role in science than the Strong Program suggests. And, like many philosophers of science, Kuhn rejected the relativism of the strong program.

Kuhn's influence in the social sciences persists, as a number of the concepts he made popular—*paradigm, incommensurability, theory-ladenness*—continue to be employed by social scientists.

K. Brad Wray

See also Kuhn on Scientific Revolutions and Incommensurability; *Naturwissenschaften* Versus *Geisteswissenschaften;* Observation and Theory-Ladenness; Paradigms of Social Science; Rule Following; Sociology of Knowledge and Science; Strong Program in the Sociology of Scientific Knowledge; Weber's *Verstehende* Approach

Further Readings

Barnes, B. (1982). *T. S. Kuhn and the social sciences.* New York, NY: Columbia University Press.

Gutting, G. (Ed.). (1980). *Paradigms and revolutions: Applications and appraisals of Thomas Kuhn's philosophy of science.* Notre Dame, IN: University of Notre Dame Press. (See especially Part II: Social Sciences)

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago, IL: University of Chicago Press. (Original work published 1962)

Kuhn, T. S. (2000). The natural and the social sciences. In T. S. Kuhn (Ed.), *The road since structure: Philosophical essays, 1970–1993, with an autobiographical interview* (pp. 216–223). Chicago, IL: University of Chicago Press. (Original work published 1991)

Martins, H. (1972). The Kuhnian "Revolution" and its implications for sociology. In T. J. Nossiter (Ed.), *Imagination and precision in the social sciences* (pp. 13–59). London, England: Faber & Faber.

Wray, K. B. (2011). *Kuhn's evolutionary social epistemology.* Cambridge, England: Cambridge University Press.

Kuhn on Scientific Revolutions and Incommensurability

It is not simple to give a concise definition of Thomas S. Kuhn's concepts of *scientific revolution* and *incommensurability* because in these somehow interwoven conceptions, all the essential elements of his philosophical position are present in a highly concentrated form. For a start, it is sufficient to characterize scientific revolutions as deep ruptures within the development of science, which separate modes of research in a particular way. The term *incommensurable* means "no common measure" and is meant to specify the relation of the succeeding traditions. The term has its origins in Ancient Greek mathematics (and particularly in Plato), where it meant no common measure between magnitudes. Since Kuhn's main work *The Structure of Scientific Revolutions* (*SSR;* 1962/1970), the metaphorical application of this mathematical notion to the relation between successive scientific theories has been a widely discussed idea. The concept of incommensurability as well as Kuhn's theory of scientific revolutions have had a great impact on the social sciences, although their applicability within this framework remains highly controversial. Nevertheless, it can be stated that Kuhn's ideas have resulted in increasing interrelations between philosophy and the social sciences.

This entry (a) gives a brief outline of Kuhn's philosophy of science, (b) explains his concepts of scientific revolution and incommensurability in detail, and (c) describes the impact of Kuhn's ideas as well as major misunderstandings of his theory.

Kuhn's Philosophy of Science

In order to attain a deeper understanding of Kuhn's concepts of scientific revolutions and incommensurability, it is necessary to sketch briefly his theory of scientific development.

Kuhn's main subject in philosophy of science is a developmental scheme, or a schematic description of scientific development, which primarily concerns the basic natural sciences, in particular physics and chemistry. The intended description is schematic in the sense that it presents scientific development as a succession of different phases whose main features are independent of the specific subject matter of the respective science. For each one of these phases a specific mode of scientific practice is characteristic. The developmental scheme that Kuhn suggests should hold true for all basic scientific disciplines, with only minor variations. It consists of the following: Before reaching maturity, nascent scientific fields are typically characterized by controversies between competing schools. There is no consensus among the practitioners of the emerging field. Each one of these schools has a particular view of the respective research domain, and typically these views are derived from extrascientific sources. Kuhn has called this phase of scientific development *preparadigmatic*. The competition among the schools may eventually come to an end when one group produces an exemplary solution to a preeminent research problem with the following two characteristics: It is sufficiently unprecedented to attract the members of the other schools, and it is sufficiently open-ended to leave enough interesting problems for further scientific work. These model solutions are called *paradigms*. They serve to implicitly guide research in the succeeding period called *normal science*. Normal science is characterized by a broad consensus by the practitioners in the field on fundamental questions, and, consequently, on a particular mode of research based on the particular paradigm.

Normal science is always confronted with *anomalies,* that is, with recalcitrant phenomena or problems that behave contrary to the expectations supplied by the paradigm. Anomalies do not usually call the validity of the guiding regulations of normal research into question. However, under special circumstances, they may, and then they are called *significant anomalies*. In such cases, the practice of science changes from normal science to *extraordinary science* or *science in crisis*. This aims to amend or even overthrow the still-binding regulations. Its research focuses on the significant anomalies and their context.

Extraordinary science resembles prenormal science in that it has a tendency to develop competing schools. However, it is more focused than prenormal science as all the various different schools must deal with the same set of significant anomalies while retaining as much as possible from the earlier period of normal science. If this research leads to a new theory that is accepted by the scientific community because it can lead to a new phase of normal science, a "scientific revolution" has occurred.

Scientific revolutions in Kuhn's *SSR* (1962/1970) are thus "the tradition-shattering complements to the tradition-bound activity of normal science" (p. 6). The rejection of the older theory is accompanied by a change in the problem field, including its related standards of solution, and by a corresponding change in basic scientific concepts. Some old concepts are discarded, some new ones are introduced, while some other concepts change meaning, sometimes in a subtle way. Kuhn even describes revolutions as *transformations of the world* in which scientific work is done, although it is not easy to make explicit and plausible what exactly is meant by this locution (something that has become a bone of contention among scholars debating the Kuhnian approach to scientific change—we shall encounter this again below). Kuhn compresses these features of revolutions into the concept of *incommensurability:* a relationship between successive traditions of normal science.

According to Kuhn, logical reasoning and empirical data play an important role in a scientific revolution, but they do not *determine* its outcome—that is, the decision for or against a new paradigm. It is rather epistemic values, for instance, the scope, simplicity, or accuracy of a theory that guide—but do not determine—theory choice during scientific revolutions. It is important to note that in this view science is basically a *social* enterprise. Scientific communities are the social enterprises that share

epistemic values and are the ultimate evaluators of knowledge claims. These communities ultimately decide between competing theories and their associated modes of research, and it is such communities that are seen by themselves and by others as responsible for some domain of scientific knowledge. Kuhn's position, therefore, has a sociological basis.

Scientific Revolutions and the Concept of Incommensurability

The most far-reaching and controversial core of Kuhn's philosophy is his thesis of the occurrence of incommensurable scientific theories as a consequence of a scientific revolution. Now, what does Kuhn mean by this notion? Let us start by considering the concept of incommensurability as it is developed in *SSR*. Incommensurability is a relational concept: It holds (or does not hold) between an A and a B. In *SSR,* A and B mainly are consecutive traditions of normal science. Think of the Ptolemaic geocentric system of the planets and the Copernican heliocentric system. In both conceptions, there was a relatively unanimous research tradition mainly dealing with the prediction of planetary positions. In such traditions that are separated by a scientific revolution, incommensurability prevails, according to Kuhn. This kind of incommensurability has three different aspects.

First, through a scientific revolution, a change in the field of scientific problems that must be addressed by any theory of that domain occurs, as well as a change in the field of problems that are legitimately addressed. Problems whose solutions are vitally important to the older tradition may disappear as obsolete or even unscientific; problems that did not exist, or whose solution was considered trivial, may gain extraordinary significance for the new tradition. Along with the problems, quite often the standards imposed upon scientifically admissible solutions change. Think of the requirement of classical mechanics—that explanations must be deterministic—which vanishes with the advent of the quantum-mechanical revolution.

A second aspect of Kuhn's (1962/1970) incommensurability theory concerns scientific methods and concepts. After a revolution, many of the older concepts and methods are still used, but in modified ways. The change of concepts discussed in *SSR* is of prime importance. It has an extensional and an intensional aspect. Extension has to do with the reference of a term (the objects it denotes or refers to), while intension has to do with the meaning of a term. The extensional aspect of a change of concepts consists of the movement of objects belonging to the extension of one concept into the extension of another concept, the two concepts being mutually exclusive. The striking example that Kuhn uses in *SSR* is the change of the concept of "planet" in the Copernican revolution. After the revolution, the earth, for instance, is a planet, whereas the sun and the moon are not planets anymore. The intentional aspect of the change of concepts consists of a change of meaning of the respective concepts. This is the case since the properties of the objects that are subsumed under these concepts change.

As the third and most fundamental aspect of incommensurability, Kuhn claims in *SSR* that, as we saw above, scientists belonging to different paradigms conduct their research in *different worlds*. But what does it mean to say that the world changes with a revolution? This question is not answered in *SSR,* although Kuhn has clearly seen its urgency there.

To understand this aspect of Kuhn's theory, it is important to distinguish between two meanings of the term *world* in Kuhn's philosophy, namely, the phenomenal world and the world in itself. It is the phenomenological world to which we actually have access in everyday life or in science. We can perceive and describe such a world, and in such a world there are octopuses, lecture halls, and electrons, for example. Such a world has a certain conceptual structure, for instance, the categories just mentioned. At this point, Kuhn holds that these concepts are of human origin; that is, we impose a structure on the world by means of these concepts, and we do not read off these concepts from the world itself, as a more familiar (realist) story would have us believe. Although it is not possible to impose any and every structure on the world, clearly more than just one is possible, which implies *a certain degree* of historical contingency of these structures. This is illustrated by the historical change of these conceptual structures (paradigms). As a consequence, paradigms are constitutive of a perceptually and conceptually subdivided world. In other words, the subjects of knowledge contribute to the constitution of the objects of knowledge (by means of paradigms) insofar as they structure the world of these objects.

The second sense of the term *world* in *SSR* is obtained by asking what is left if one subtracts all these human (subject-sided) contributions—that is, when we subtract all this perceptual and conceptual structuring from the world taken in the first sense. Then one is left with a world that is completely independent of our perceptions and conceptions, a world—as one might say—that is purely object-sided. But we have, according to Kuhn, no access whatsoever to this purely object-sided world. According to Kuhn, what we, or a particular scientific community, describe as reality is *a*—but not the only possible—phenomenal world. Due to the historical change of these worlds, Kuhn's theory has sometimes been described as "Kant on wheels."

Furthermore, because of such incommensurability, we must, according to Kuhn, rethink the concept of *scientific progress* in the natural sciences. First, scientific progress is not cumulative, due to conceptual changes during revolutions. Cumulativity implies that something that is a part of science at some point in time will, neglecting small corrections, remain a part of science forever. However, the conceptual changes that occur during revolutions are much more than small corrections of the existing body of knowledge. They amount to a thorough conceptual reorganization, and typically to a changed ontological perspective. Furthermore, Kuhn denies that scientific progress is an approach to truth. Instead of conceiving scientific progress as a goal-directed process, we should think of scientific progress in an analogous way as Darwinian evolutionary theory conceives of evolution. Darwinian evolutionary theory states that there is no goal toward which evolution is directed. In a similar way, in scientific development, there is no "set goal," which would be a "permanent fixed scientific truth," that science approaches.

In *SSR,* the concept of incommensurability is not entirely clear. Thus, it has been the subject of much criticism as well as misunderstanding. Most of Kuhn's later work has further developed and refined the conception of incommensurability in order to answer critical questions regarding his ideas of "world change" and "semantic incommensurability." For instance, is it really appropriate to speak of the Kuhnian cases as examples of meaning change? Is it not that the respective concepts themselves remain constant and are only used differently? And would it even speak against "approaching the truth" if the concepts actually did change in meaning, as long as they still refer to the same entities as before, but in a more precise way (as a scientific realist might ask)? It is mainly these questions and problems that Kuhn worked on from the late 1960s on. However, his answers to these questions remain controversial.

The Impact of Kuhn's Ideas and Major Misunderstandings

Kuhn's theory brought forth a significant change within the philosophy of science of the 20th century as it stressed the need to include sociological thinking and in particular detailed knowledge of the actual history of science in order to arrive at a reasonable philosophy of science. This post-positivist turn was antithetical to the "standard conception" of science as advocated by logical positivists as well as Karl Popper and led to the emergence of various new subdisciplines in philosophy and the social sciences (and between them), such as studies in history and philosophy of science, science and technology studies, and the developing field of social epistemology. Furthermore, Kuhn's ideas were widely adopted by the social sciences. The predominant use of Kuhn's work in the social sciences concerns the current status of a particular discipline or its history. Is the discipline still in its preparadigmatic state, or does it already have a paradigm? Are the changes that the discipline experienced in the past veritable Kuhnian revolutions? Could competing theories, say in sociology, be described as incommensurable? These questions received the most diverse answers from different authors. Elements from Kuhn's theory were used by some to defend mainstream social science and by others to develop a critical attitude toward it or to defend or attack alternatives. Some authors attempted to show that on Kuhnian terms, a given social science is very similar to natural science, whereas others tried to demonstrate the opposite, and so on. No consensus has been reached in any of the social sciences about these and related questions. Kuhn himself said very little about the social sciences, basically only that they are typically in the preparadigmatic state. This, however, did not prevent many social scientists from uncritically applying parts of Kuhn's theory to different aspects of their disciplines, without further arguments or corroborating case studies. On top of that, Kuhn's ideas have been (mis)used for the most diverse purposes. On the basis of his—indeed sometimes ambiguous—texts, even outright contradictory assertions have been

made by some social scientists, often to promote their own agenda. Take, for example, the sociology of science following Kuhn: With the advent of *SSR,* the production of scientific knowledge seemed not at all determined by time-independent, universal methodological rules anymore. Instead of binding rules, research and theory choice in times of extraordinary science is guided by values that only influence but do not determine scientific behavior. Such values are influenced by group-specific and even idiosyncratic factors. In other words, scientific knowledge now indeed seemed socially contingent, which licenses and even appears to demand sociological analysis. Typically, the content of science was now seen as the result of negotiations among scientists in which human interests of a personal or political nature, or power relations, played an important and sometimes even decisive role. Science, therefore, was assumed to be a completely contingent enterprise without any rational aspects. It should be noted, however, that Kuhn himself remained extremely skeptical toward those claims. In his view, the *epistemological* dimension of scientific knowledge—the possibility of its normative evaluation from within science—must be an essential part of an integrated image of science, but it had been entirely dismissed by many social scientists.

Next to this point, Kuhn's concept of incommensurability has often been understood as the "absolute incomparableness" of scientific theories with respect to their scientific merits, implying that it is impossible to rationally compare and evaluate successive scientific theories. At first glance, this seems to be a compelling consequence of incommensurability. Since incommensurable theories deal with "different worlds," they seem to stand in the same relationship as theories about the unconscious stand to theories about the stability of galaxies: These theories clearly describe different domains with mutually untranslatable vocabularies. Because there cannot be any empirical friction between such theories, there cannot be any real competition between them, and consequently, the question of a rational choice between them cannot arise either. The same seems to hold true for incommensurable theories.

However, that is not Kuhn's point at all. Instead, he asserts that it is impossible to compare two theories point by point; that is, he denies the possibility that each (general) statement of one theory (e.g., Newtonian mechanics) can be confronted with the corresponding statement of the other theory (e.g., Einstein's theory of relativity). This holds since, with the new conceptual vocabulary, statements can be formulated that are incapable of articulation by the old vocabulary. Additionally, even corresponding empirical statements may carry different weights among the two theories, so that the merits of one theory may be depreciated from the viewpoint of the other. Yet the two theories can be compared globally, with respect to their simplicity, accuracy, fruitfulness, predictive power, and so on. Nevertheless, two scientists in agreement on the list of such epistemic values may still disagree on which theory should be preferred, for example, due to different weights of the shared values. This is indeed the case during the phase of extraordinary science. Therefore, theory comparison and theory choice do not resemble an algorithmic procedure in which all applicants of the algorithm must get the same since they have to follow fixed rules mechanically. Yet this form of theory comparison is far from being irrational, at least if one supposes that the sort of epistemic values mentioned form a reasonable ground for theory choice within scientific communities.

Paul Hoyningen-Huene and Simon Lohse

See also Feyerabend, Critique of Rationality in Science; Kuhn and Social Science; Lakatos, Methodology of Scientific Research Programs; Paradigms of Social Science; Popper's Philosophy of Science; Relativism in Scientific Theories; Social Epistemology; Social Studies of Science and Technology; Sociology of Knowledge and Science

Further Readings

Bird, A. (2000). *Thomas Kuhn.* Princeton, NJ: Princeton University Press.

Hoyningen-Huene, P. (1993). *Reconstructing scientific revolutions* (A. T. Levine, Trans.). Chicago, IL: University of Chicago Press.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago, IL: University of Chicago Press. (Original work published 1962)

Kuhn, T. S. (1977). *The essential tension.* Chicago, IL: University of Chicago Press.

Kuhn, T. S. (2000). *The road since structure* (J. Conant & J. Haugeland, Eds.). Chicago, IL: University of Chicago Press.

Preston, J. (2008). *Kuhn's "The Structure of Scientific Revolutions."* New York, NY: Continuum.

Encyclopedia of *Philosophy* and the *Social Sciences*

Editorial Board

Encyclopedia of *Philosophy* and the *Social Sciences*

2

Byron Kaldis

The Hellenic Open University, Greece

Editor

SAGE reference

Los Angeles | London | New Delhi
Singapore | Washington DC

Los Angeles | London | New Delhi
Singapore | Washington DC

FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London, EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Printed in the United States of America.

A catalog record of this book is available from the Library of Congress.

ISBN 978-1-4129-8689-2 (cloth)

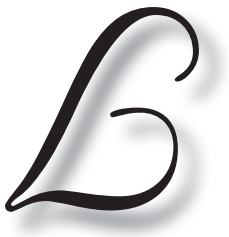
Publisher: Rolf A. Janke
Acquisitions Editor: Jim Brace-Thompson
Developmental Editors: Diana E. Axelsen, Sanford Robinson
Production Editor: David C. Felts
Reference Systems Manager: Leticia Gutierrez
Reference Systems Coordinator: Anna Villasenor
Copy Editor: QuADS
Typesetter: Hurix Systems Pvt Ltd.
Proofreaders: Jeff Bryant, Christine Dahlin, Pam Suwinsky
Indexer: Virgil Diodato
Cover Designer: Candice Harman
Marketing Manager: Carmel Schrire

13 14 15 16 17 10 9 8 7 6 5 4 3 2 1

Contents

Volume 2

List of Entries



Lakatos, Methodology of Scientific Research Programs

Imre Lakatos (1922–1974) was an émigré who left his native Hungary for Britain during the 1956 uprising. He had already completed a doctorate from the University of Debrecen, having submitted a dissertation titled *On the Sociology of Concept Formation in the Sciences*. A further doctorate was completed at the University of Cambridge, with the thesis *Essays in the Logic of Mathematical Discovery*. In 1960, he became a member of the Department of Philosophy, Logic, and Scientific Method at the London School of Economics, eventually becoming a professor in 1969. He died suddenly at the age of 52. The three major influences on Lakatos were Karl Popper, the mathematician George Pólya (with his heuristic approach to mathematics), and in a strange way the Marxist György Lukács (since Lakatos's approach to matters can be broadly called "dialectical").

As the titles of his two doctorates indicate, his overall main interest was in the *dynamical growth* of theories, both within mathematics and the empirical sciences, and in the rationality that such growth exhibits. The emphasis on growth put him at odds with more formal and axiomatic approaches in science and mathematics, and this took him in the direction of the history of the development of ideas in these fields. However, Lakatos did not regard himself strictly as a historian, since he provided what he called *rational reconstructions* of developments in mathematics and the sciences, which may well be at odds with the fine details of the actual course of the development of the sciences. Lakatos's approach gives us deep insights into the complexities of the historical growth in all sciences, from mathematics to the social sciences (including economics), and the ways in which the growth can be deemed to be rationally based.

The Growth of Science I: Proofs and Refutations (Deductive Science: Mathematics)

Lakatos's first major publication, a sequence of four articles in the *British Journal for the Philosophy of Science* 1963–1964, was drawn from his Cambridge thesis and titled *Proofs and Refutations* (it is dedicated to Pólya and Popper). At the time of his death, revisions were not complete; but essentially, the original papers along with other writings were published in the 1976 collection brought out by Cambridge University Press, with other papers on philosophy of mathematics appearing in the 1978 publication *Mathematics, Science and Epistemology* (Philosophical Papers, Volume 2). It is cast in the form of a dialogue between a teacher and pupils and investigates the conjecture that for all polyhedra the number of *V*ertices minus the number of *E*dges plus the number of *F*aces is equal to 2, namely, $V - E + F = 2$ (a result originally proposed by Leonhard Euler). In the dialogue, a "proof" of this is offered; it is not a formal proof in the sense of mathematical philosophers like Gottlob Frege, Bertrand Russell, or David Hilbert but rather an informal proof that depends on a number of "thought experiments" that are embedded in other subconjectures or lemmas, such

as whether solids can be "projected" into deformable rubber sheets.

As the dialogue amply shows, the proof is open to "counterexamples" or *refutations*. What should one do in the face of alleged counterexamples? Either one could accept them, in which case Lakatos elaborates a "method of surrender," or one could reject them, in which case Lakatos elaborates a "method of monster barring." The latter approach may be such that if the Euler theorem does not hold of a figure then it can be denied that the figure is really a polyhedron. This in turn leads to the strategy of *concept contraction* or concept stretching. In a brief outline, this is Lakatos's more dialectical approach involving a "method of proof and refutations," which, he insists, really underlies the way in which axioms and proofs in mathematics are discovered and then critically evaluated, something obscured by more formal approaches.

Later in the dialogue, Lakatos proposes four heuristic rules of procedure. But these are not fully developed, nor do they provide a recipe for generating proof or refutations. Rather, they are guidelines that govern the way in which hypotheses within a theory evolve as a sequence of proofs and refutations. Just how closely does such a procedure capture the actual history of what has historically gone on in mathematical thought in this area? Lakatos (1976) says that it "should reflect the dialectic of the story; it is meant to contain a sort of *rationally constructed or 'distilled' history. The real history will chime in the footnotes . . . as an organic part of the essay*" (p. 5). Lakatos's approach generated a lot of interest at the time in the philosophy of mathematics, but few followed him fully in rejecting what he dubbed the "formalist" metamathematical approach to mathematics, which in his view "disconnects the history of mathematics from the philosophy of mathematics" (p. 1).

The Growth of Science II: Scientific Research Programs (Inductive Reasoning: The Empirical Sciences)

In 1965, Lakatos organized at the London School of Economics an international colloquium in the philosophy of science, which attracted leading philosophers of science from around the world. The proceedings were published in four volumes, with Lakatos himself contributing some important papers, two of which will be mentioned. The first of these is a long paper analyzing the approach of Popper and Carnap to the problem of evidence in science, titled "Changes in the Problem of Inductive Logic" (reprinted in Lakatos, 1978b, chap. 8). Lakatos had picked up on a phrase previously used by Popper—"research program." Lakatos developed this as a framework within which problems can be posed and solved and that may be characterized as "progressive" or "degenerating." Framing matters this way, Lakatos was able to present an important part of the history of the development of inductive reasoning in a manner akin to the way he had done for proofs and conjectures within mathematics.

Even more important is Lakatos's "Falsification and the Methodology of Scientific Research Programmes" (SRPs), in which he turns to the application of his general approach to the empirical sciences. This originally appeared in Lakatos and Musgrave (1970) (reprinted in Lakatos, 1978a, chap. 1). The conception of science and scientific method that Lakatos proposed has had wide application not only in the physical sciences but also in the social sciences, particularly economics.

As in his work on mathematics, Lakatos's framework focuses on two aspects of science: (1) the dynamical growth of theories and (2) the rationality that this growth exhibits. In respect of (1), Lakatos's position resembles that of Kuhn's paradigms, but the emphasis on (2) gives a much more prominent role to the rationality of science. Also, his framework takes him away from the common axiomatic approach to the structure of scientific theories (something that is also characteristic of his approach to mathematics, as shown above). Since few theories are presented in an axiomatized form and even fewer are given a more rigorous formal presentation, Lakatos's theory of SRPs, with its more historically presented dynamical structure of theories, was welcomed.

Lakatos proposed that SRPs be understood as a historically generated (finite) sequence of theories (T_1, T_2, T_3, . . .) each of which is a conjunction of a "hard core" (HC) and a "protective belt" of auxiliary hypotheses A_i (i.e., T_i = (HC) & A_i). The HC is a set of hypotheses that remains the same for the life of the program. It need not be true, but it is treated as if it were true by those working on the program. The auxiliaries are the variants, which can change from theory to theory within the program. These may be simply additional hypotheses, or they could

be descriptions of models to which the hypotheses of HC are to apply. Lakatos illustrates his approach in the case of Newton's theory. In the Newtonian SRP, the HC is Newton's three laws of motion and the law of universal gravitation. The auxiliaries are the various models of, say, the solar system, which is initially understood as a single-point particle orbiting a central-point particle; this is then developed to remove simplifications so that the two particles become massive bodies; they are then (in the model) made to rotate on their axes; they are then embedded in an even more realistic model of the solar system in which a number of planets orbit a central body each of which perturbs the other, and so on.

A second important feature of an SRP is the heuristic associated with it. The *negative heuristic* bids that when any of the sequence of theories is in trouble, the HC is not to be altered, but the auxiliaries are to be changed. The *positive heuristic* specifies guidelines to which alterations to the auxiliaries are to conform; but they do not provide an "algorithm" that tells us *how* to alter the auxiliaries. So in the Newtonian SRP, the negative heuristic does not permit us to change Newton's fundamental laws, while the positive heuristic tells us to treat all systems as mutually gravitationally attracting bodies. The heuristic may be sufficiently powerful for the program to develop under its own steam and uncover many novel facts without necessarily focusing on the anomalies that it might face. As Lakatos says, a program can develop in an "ocean of anomalies."

Rationality in Scientific Growth

The *rationality* of a program can be determined by investigating its sequences of theories. A pair of theories may have the following characteristics. The second is *theoretically progressive* if it has consequences (e.g., predictions) that the first does not. And it will be *empirically progressive* if some of those consequences turn out to be true—that is, a new fact is predicted (if none are true, then it is *empirically degenerating,* i.e., Lakatos says it is *ad hoc*$_2$). A pair will be *theoretically degenerating* if the second makes no predictive advance over the first; and it will be *theoretically stagnating* if it makes even fewer predictions (i.e., it is *ad hoc*$_1$). Finally, if the SRP runs out of heuristic steam, it might well accommodate facts relevant to it but uncovered in a rival program by simply altering the auxiliaries in any fashion (perhaps breaking the positive heuristic); it is then said to be *heuristically degenerate* (i.e., *ad hoc*$_3$). This leads to a *criterion for demarcating science from pseudoscience:* An SRP must have at least one phase that is empirically progressive.

Assessing Rival Scientific Methodologies

When Lakatos developed his methodology of SRP (MSRP), it led to a number of comparative studies about how various methodologies, such as inductivism, probabilism, Popperian falsificationism, and conventionalism, can deal with some of the central episodes in science; and it also invited comparison with methodologies that tended to be nonrational from Thomas Kuhn's theory of paradigm shifts (though the later Kuhn made more explicit his methodology of weighted values), Polanyi's account of tacit knowledge, and Paul Feyerabend's antimethodology stance.

Lakatos proposed a metacriterion for judging rival methodologies: We are to adopt that methodology that maximizes rational explanations of moves in the game of science and minimizes other moves, which are then relegated to leftovers for sociologists of science to explain. Lakatos and his supporters argued that MSRP gave maximally the best explanations. This led to a wide-ranging critical discussion of the value of various theories of method and complaints from historians of science that MSRP was a distorting lens through which to view the complexities of actual science. But as in the case of his approach to the history of mathematics, Lakatos took his application of MSRP to the sciences to give a *rational reconstruction* of that history and not one that gets lost in the complex details of actual history. For Lakatos, methodology without history of science is empty, but history of science without a rationalizing methodology is blind.

Robert Nola

See also Explanation, Theories of; Falsifiability; Feyerabend, Critique of Rationality in Science; Kuhn on Scientific Revolutions and Incommensurability; Popper's Philosophy of Science; Pseudoscience; Scientific Method

Further Readings

Howson, C. (Ed.). (1976). *Method and appraisal in the physical sciences: The critical background to modern*

science 1800–1905. Cambridge, England: Cambridge University Press.
Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery.* Cambridge, England: Cambridge University Press.
Lakatos, I. (1978a). *Mathematics, science and epistemology* (Philosophical Papers, Vol. 2; J. Worrall & G. Currie, Eds.). Cambridge, England: Cambridge University Press.
Lakatos, I. (1978b). *The methodology of scientific research programmes.* (Philosophical Papers, Vol. 1; J. Worrall & G. Currie, Eds.). Cambridge, England: Cambridge University Press.
Lakatos, I., & Musgrave, A. (Eds.). (1970). *Criticism and the growth of knowledge.* Cambridge, England: Cambridge University Press.
Latsis, S. (Ed.). (1976). *Method and appraisal in economics.* Cambridge, England: Cambridge University Press.

Language, Philosophy of

The philosophy of language has played an extremely important role in 20th-century and early-21st-century philosophical thought, but it is no newcomer to the scene. Plato's dialogue *Cratylus* is a sustained look at the nature of word meanings that contrasts conventionalism about meaning with naturalism. Aristotle developed a logic of quantification and of modality and showed that we can study both natural language and logic using a formal language. The Stoics, too, made contributions to linguistic analysis, provided a division of the parts of language into five categories, and offered an analysis of these parts based on syntactic and semantic functions. They also contributed to the notion of complete and incomplete meanings, thus contributing to the theory of propositions and the semantics of predicates. Work on vagueness, ambiguity, and logic continued through the medieval and modern periods. However, the modern watershed achievement is the invention of modern logic by the great mathematician, logician, and philosopher Gottlob Frege and his subsequent work in both logic and metalogic. Frege's generalization of the mathematical notions of function and argument and his logic's ability to handle the interaction of multiple quantifiers provided the impetus for the "linguistic turn" that has characterized much of 20th-century analytic philosophy. This turn has resulted in the vigorous study of various logics (modal, temporal, erotetic, deontic, etc.). It provides analytic philosophy with its focus on logic and language and reveals how the exploration of these informs philosophical questions in general.

A few of the puzzles considered within the bailiwick of philosophy of language are as follows:

- What is meaning, and how do words, phrases, and sentences acquire it?
- What logic (if any) governs natural language?
- Does the language you acquire affect your cognitive and conceptualizing abilities?
- What are the different actions (especially speech acts) we can undertake using language?
- What norms govern language use?
- Are all philosophical puzzles really linguistic puzzles?

These questions are obviously related. Generally, the philosophical study of language has been divided into three parts: (1) syntax, (2) semantics, and (3) pragmatics. *Syntax* is the study of the rules and principles that generate the sentences and phrases in a language out of its basic parts. *Semantics* is the study of (linguistic) meaning. *Pragmatics* is the study of language use. Many linguistic phenomena are difficult to classify as belonging properly to any one area, and the borders of semantics and pragmatics are often highly controversial.

Syntax

One very popular approach to syntax was initiated by Noam Chomsky, leading to a conception of grammar as (a) universal, (b) known innately by humans, (c) generative, and (d) the study of an individual's linguistic competence—her I-language (or internal language), as opposed to E-language (or external language, e.g., French or English). The features of grammar (a) to (d) are interrelated: Universality is the result of generative rules that humans know as part of their linguistic competence.

The focus of these sorts of theories has raised very interesting questions in epistemology. In particular, there is controversy over the nature of the epistemic relation humans have to the rules of generative grammar that characterize their linguistic competence. Chomskians argue that it is simply propositional knowledge of various parameters that constrain possible grammars (albeit knowledge that the users find extremely difficult to articulate), but

others have found this sort of tacit knowledge to be extremely puzzling. The innateness of the knowledge is controversial as well. It is often justified by an argument from the *poverty of the stimulus:* Children aren't exposed to enough evidence, namely, external stimuli, in order to explain the rapidity and ease with which they master a language.

Philosophers often proceed on the assumption that syntactic representations are the inputs for semantic interpretation. This requires reification of the objects of syntactic theory as things to be interpreted. This reification, in many syntactic theories, requires positing a great deal of "unvoiced structure"—parts of the representation that receive no actual phonological execution at all. For example, it is often argued that "to leave" in (1) contains an unvoiced subject as represented by (2):

1. John wants to leave.
2. $John_1$ wants [PRO_1 to leave]

"PRO"[1] is co-indexed with "John," and semantically, it is interpreted as coreferring with John (if "John" refers at all). This is a surprising result but not one without justification. If PRO is represented in (1)'s LF,[2] it is fair game for semantic interpretation. The extent to which a sentence's syntactic structure constrains its meaning in a context is a hotly debated issue.

Semantics

Two constraints that seem to apply to any theory of linguistic meaning are as follows:

1. *Compositionality (C):* The meaning(s) of complex linguistic units are determined by the meanings of their parts and how the parts are syntactically structured.
2. *Publicity (P):* Linguistic meaning must be public.

(C) is not obviously true of all meaningful units: The idiomatic meaning of "kick the bucket" doesn't seem to be a function of the meanings of the words *kick, the,* and *bucket.* However, linguistic meaning is generally argued to be compositional since we can understand novel sentences. Consider the following:

(NS): A gentleman from Nova Scotia who travels with a fish was accosted by a troop of extremist ballerina protestors.

(NS) has a clear meaning despite its novelty. This is difficult to explain on any theory that doesn't explain the structure of complex units via the meanings of their parts and how they are structured. (P) is plausible because we can communicate using language, which would be inexplicable if meaning was private. On this basis, Frege argues that communicable meanings can't be subjective ideas. Some philosophers (cf. Wittgenstein) argue that private meaning is impossible (and not merely incommunicable). Like most bold claims in the philosophy of language, the extent to which privacy is compatible with meaning has proven highly controversial.

If (communicable) meaning has to be both compositional and public, it is natural to think of meaning as involving something that linguistic users can all converge on. Several options have been advanced as possibilities that can be grouped together. In general, philosophers of language have taken indicative sentences as a starting point. This isn't surprising given the assumption that language is a device of communication and information exchange. This leaves imperatives, interrogatives, performatives, and the like on the side, since "Finish your spinach" isn't truth evaluable. Philosophers of language have generally hoped that an account of the meaning-indicative sentences will provide the basis for a theory of the other parts of speech. "Finish your spinach," for example, may have satisfaction conditions involving states of affairs in which "You finished your spinach" is true.

Meaning and Reference

If meaning is compositional, the atomic parts of sentences must have meanings. It has proven useful to posit a relation—*reference*—that holds between atomic units of meaning and the world. While it is fairly easy to make sense of names referring, the thesis that predicates, quantifiers, aspect, and tense morphemes refer is much more controversial.

If we posit reference for predicates and names, we can offer a toy theory of the meaning for the following:

(A) Agamemnon fathered Orestes.

(A)'s meaning (ignoring tense) can be given as a composition of the referents of *Agamemnon, Orestes,* and *fathered.* We can then deal with the tense marker by treating it either as an operator on the sentence, which effectively means "in the past of the time of the context of utterance of (A)," or as

referring to a time in the past. The meaning of (A) is thus compositionally determined by the referents of the parts of the sentence and how it is structured.

The theory faces two well-known problems. First is the problem of empty names. Neither "Zeus" nor "Pegasus" refer, but they seem to contribute different meanings to sentences. Second is the problem of informative identities. While

(SS) Spiderman is Spiderman is trivially true,

(SP) Spiderman is Peter Parker

is true but highly informative. Mary Jane can believe (SS) without believing (SP). However, given that Spiderman and Peter Parker corefer, compositionality requires that (SS) and (SP) mean the same thing.

A solution, according to Frege, divides meaning into two dimensions: (1) sense and (2) reference. The role of sense is to determine reference, and senses can bear a many-to-one relation to the referent they determine. Thus, while Peter Parker and Spiderman corefer, the sense of "Peter Parker" is different from the sense of "Spiderman." (SS) and (SP) express different compositionally determined semantic objects at the level of sense, but they determine the same reference. We can explain both why (SP) is informative and why one can believe (SP) but not (SS). However, positing senses for names is the target of many powerful arguments that have led philosophers to rethink the relationship between informativeness and meaning. It is generally an open question as to how the cognitive aspects of reference are to be integrated into a theory of meaning.

Another interesting set of questions regarding reference involves context-sensitive expressions. Indexicals (*I, here, now*) and demonstratives (*this, that*) have stable meanings but refer to different objects on different occasions of use. Providing a semantics for indexical expressions led to a view that sentence truth is doubly indexed—to both a context of use and the circumstances relative to which you evaluate the sentence. This approach opens new venues for the study of context and meaning. For example, a popular view in epistemology takes "know" to operate in a manner similar to "now": It has different referents in different contexts of utterance.

Sentence Meaning

Some sentences in different languages are synonymous, leading to the question of whether or not to reify sentence meaning. Those who are happy to do so think of meanings as propositions, though there is little consensus on what propositions are. On one view, the meaning of a sentence is the set of possible worlds at which it is true: Propositions are neither true nor false, but they are true or false at worlds. This view of propositions is beset with a challenge: Surely, "2 + 2 = 4" and "3 + 3 = 6" don't mean the same thing even though they are true at all the same worlds (i.e., all of them). A more fine-grained view of propositions takes propositions to be set theoretic structures containing either objects, properties, and relations as members, or senses. Both views face the problem of the unity of the proposition: What unifies these structures in the right sort of way to constitute something thinkable, judgeable, and truth evaluable?

Philosophers eschew reification of sentence meaning and insist that meaning is to be identified with truth conditions, in a theory that specifies "T sentences" such as

S is true iff p.

S ranges over (structural descriptions of) sentences of the language we are giving the theory for (the "object language"); the rest of the sentence is in the language we are using to give the theory (the "metalanguage"). If English is used as a metalanguage for French, we get

(F) "La neige est blanche" is true iff snow is white.

(F) is informative and displays what the object-level sentence means. One prominent worry for this sort of theory is that the "iff" is an extensional connective: It's true only in cases where the two sentences it conjoins have the same truth-value. Consider (G):

(G) "La neige est blanche" is true iff snow is cold.

(G) is true but not a suitable candidate for specifying the meaning of "La neige est blanche." Truth-conditional theorists try to solve this by invoking *radical interpretation* and *charity:* You take your subjects and interpret their language by trying to make most of what they say (and believe) come true and be reasonable. While the success of this gambit is much disputed, the idea that some principle of charity constrains interpretation has been very powerful in the philosophy of language.

Other theories tie meaning to epistemic abilities. Logical positivists tried to explain sentence meaning in terms of verification (or falsification) conditions, while ordinary language philosophers took sentence

meaning to be the set of sanctioned moves in a language-game that the community both played and policed. Both views had far-reaching implications for philosophy, seeing most traditional philosophical problems as simply the result of a subtle misunderstanding of meaning.

Externalism and Internalism

The philosophy of mind and the philosophy of language share an interesting question: Is the content of a thought or sentence determined by the psychological properties of the thinker/speaker or by external features (as well)? Powerful arguments have been marshaled to show that the environment plays a role in determining content. Often these arguments take the following form: Imagine a perfect duplicate of yours who uses the word *w* in the same way you do but for whom the apparent referent of *w* is different in microstructure from the actual referent of *w*. Intuition has been taken to favor the view that *w* is distinct in referent and the speakers distinct in their thoughts as regards *w*.

Externalism is implicated in another set of arguments concerning meaning and underdetermination. Take any set of facts that can plausibly serve the role of determining meaning. Any number of rules regarding the use of these terms can be "fitted" to these facts. In fact, results in metalogic show that we can take any theory we like and interpret it using only the natural numbers and retain all the same theorems. This means that unless the world provides better or worse referents for our terms it is very hard to see what could possibly ensure that words refer to the referents we take and expect them to refer to. Some have taken this result to be a clear victory for anti-realism and pragmatism. Others have taken this result to show that the world really does come with better and worse referents for our terms—some properties (or classes) are simply more natural than others, and naturalness makes for more eligible candidates for reference.

Pragmatics

Pragmatics is broadly divided into two parts. First, there is *speech act theory*. We use language in all sorts of interesting ways: We assert, we question, we command, we flirt, we apologize. Often the same sentence can be used to perform different speech acts: "The police are coming" could be an assertion or a warning. While the study of assertion has dominated much of the literature, there is lively study of the other types of speech acts we perform and how they interact with the meaning of the sentences we use to perform them. The study of speech acts connects closely with the study of mind: Many of the felicity conditions governing speech acts depend on the speaker being in correlated mental states. Assertion is arguably felicitous only if the speaker believes the content of the assertion. Promises aren't felicitous without the correlative intention to make the content of the promise true.

Second, we often use sentences in order to convey more information than they literally mean. Responding, "I am free tonight" to a request to go out for drinks *implies* that you will accompany your interlocutor without *saying* so. This relation is not accidental, and philosophers of language have appealed to the rational, cooperative character of conversation to explain these implications, which are known as "implicatures." A complicated issue arises over the scope of these sorts of explanations. If Joe claims that "it is raining" when in Paris, we can work out that he means that it is raining in Paris, not merely that it is raining. If Joe claims that "Tipper didn't quit smoking," we can work out that Tipper used to smoke even though Joe didn't assert that she did. Accounting for these cases under the same rubric as implicature has been and continues to be a matter of much controversy.

Language and Thought

There are a great deal of interesting questions about the interaction of language and thought. One extreme claim is that an individual's language has very strong effects on her cognitive and conceptual abilities, though it isn't clear how to make this sort of claim more precise and testable. There are powerful philosophical arguments to the effect that linguistic relativity of any strong sort is literally a priori incoherent. The universality of various domains of thought remains an interesting and controversial issue.

A related, bold claim is that thought itself is done using a language of thought (LOT). LOT theorists claim that a naturalistic, computational theory of mind presupposes a LOT given certain apparent facts about the compositionality of thought. The thesis is controversial as it requires that we have a language that we don't learn but simply possess and whose linguistic parts are, at least in part, not learned but simply had.

Finally, there is broad agreement that what a speaker means by uttering a sentence has a lot to do with the speaker's intentions in uttering it. An ambitious project has involved trying to give as large a role as possible to a speaker's intentions in determining meaning in general in order to ground meaning in the operations of mind. This approach usually involves a speaker's intentions in both exchanging information by her utterance and having her interlocutor bare complex relationships to the information in recognition of the speaker's intention. The success of this project, however, is a difficult and controversial issue.

Adam Sennet

See also Analytic/Synthetic Distinction; Concepts; Holism, in the Philosophy of Language; Language and Society; Language-Games and Forms of Life; Logical Positivism/Logical Empiricism; Rule Following; Semantics and Pragmatics; Social Anti-Individualism and the Mental; Speech Acts; Truth, Philosophical Theories of; Verificationism

Notes

1. PRO is, roughly speaking a pronoun in generative grammar that provides a subject to clauses such as "to stay."
2. LF is a level of syntactic representation that is often taken to be the structure that is interpreted semantically. It plays the role of providing the sentence's logical and semantic structure.

Further Readings

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York, NY: Praeger.

Davidson, D. (1967). Truth and meaning. *Synthese, 17,* 304–323.

Frege, G. (1960). On sense and reference. In P. Geach & M. Black (Eds. & Trans.), *Translations from the philosophical writings of Gottlob Frege* (p. 58). Oxford, England: Blackwell.

Grice, H. P. (1989). *Studies in the ways of words*. Cambridge, MA: Harvard University Press.

Kaplan, D. (1989a). Afterthoughts. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 565–614). Oxford, England: Oxford University Press.

Kaplan, D. (1989b). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–564). Oxford, England: Oxford University Press.

Kripke, S. (1980). *Naming and necessity.* Cambridge, MA: Harvard University Press.

Russell, B. (2005). On denoting. *Mind, 114*(456), 873–887.

Stalnaker, B. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics* (Vol. 9, pp. 315–332). New York, NY: New York Academic Press.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford, England: Blackwell.

Language and Society

Society consists entirely of interactions between people, to the extent that people discuss them. Put differently, society is the ensemble of social structures to which meaning is assigned. Society can be viewed as the verbalized conventions about how to interact with each other, including ways to distinguish between "us" and "them," all the stories told, all the "facts," and all the ideas, including all the controversies talked about, everything said about the visible and the invisible worlds, and the entirety of arguments exchanged by its members.

In order to discuss the relationship obtaining between discourse and society, we therefore have to look at the role of discourse in the construction of reality, social reality in particular. What distinguishes (human) society from a horde of chimpanzees is that one part of the social structure, namely, all that is symbolic, is discursively negotiated between its members. There are no doubt behavioral universals embedded in our genes, and also there is much interpersonal behavior acquired by observing and imitating other members of the group. While the acceptance of hierarchy may be programmed by our genome, the ways we express submission may to some extent be learned. But without engaging in discourse, we cannot assign meaning to the behavior we encounter or form any ideas, for instance, the idea of hierarchy and, its inverse, the idea of equality.

Society and its foundation—that is, linguistic discourse—are therefore what distinguish groups of humans involved with each other from chimpanzees. Chimpanzees also live together in social structures. They interact with each other in ways we often find similar to our own ways. Mothers tend to their offspring, the young ones learn lots of things by imitating grown-ups, and grooming maintains interpersonal relationships. They share food, they cooperate when attacked, and they seem to organize

themselves in a hierarchy. Yet these social structures are defined by nature, not by culture. There is no "signifying system"—no signs standing for something else—that would enable them to communicate symbolic content. They have no control over their social structures. Groups of humans, on the other hand, are defined both by nature and by culture. Often it is not easy or even possible to clearly draw a line between the two. Language, however, is one such line. Language is symbolic, nature is not.

Human society organizes itself along the ideas exchanged between and shared by people. Chimpanzees do not talk about their social structures. Like humans, they distinguish between "us" and "them," but they are not known to enter an argument on who is friend and who is foe. Humans do. They keep discussing, for instance, whether group membership is based on territory, bloodline, race, or class. Concepts such as *race* or *bloodline* are discourse constructs, and they mean what people contributing to the discourse say they mean. They refer to nothing in particular in the world outside. Such concepts have their place only in the discourse, not in a discourse-external reality. Thus, language replaces instinctive behavior by a behavior based on shared ideas, a behavior based on culture, whether individually accepted or not. Chimpanzees do not discuss their mates' behavior, while humans continually talk about it, assigning meaning to it. Chimpanzees behave without being aware of their behavior; they do not carry out acts, and they are not known to talk about their intentions. Humans, on the other hand, have learned to explain why they do what they do. To have ideas means also to be aware of them. This happens when people tell us about them. Once we have become accustomed to living in a world of ideas, we replace the world out there, the world that has no meaning, by a world to which meaning has been assigned through language. To a large extent, we have substituted culture for nature.

Language, as understood here, is not a fixed system, a kind of innate language organ humans are endowed with; it is what people say (or sign or write) and what they recognize as being said. Language is communication by signs. Language thus is social, and this aspect of language is what is here called *discourse:* It is the entirety of utterances that have been and are being communicated between people ever since they began talking to each other. Discourse comes with its rules and conventions. They, too, are not inherent in a language system; they are worked out by speakers. As long as people keep talking, discourse will grow, and while it grows, it will undergo gradual change. While all human society shares the capability for language, each group, separated from other groups by the specifics of their discourse, develops its own culture. Culture is about sharing ideas and thus presupposes discourse. Culture is about negotiating distinctions. Therefore, we should talk about different discourse communities rather than about different societies.

There are as many discourses as there are discourse communities. Whenever they feel like it, people can form new discourse communities. What defines a discourse community is how it distinguishes itself from other communities. Though every individual owes her or his symbolic world to discourse, each has her or his own world, different from anyone else's, due to the interactions in which she or he took part, due to the discourses she or he took part in, and of course also due to the ways in which she or he has processed the content of these interactions, forgetting much, remembering a lot wrongly, and recalling very little correctly. It makes sense to look at a discourse as the collective mind of the apposite discourse community. But being part of a collective mind does not imply sameness of the individual minds of which it is made up.

Discourse and the Construction of Reality

Discourses, and discourse communities, are cultural constructs. They are not "real"-world entities.

Whenever humans are introduced to a new thing, they are taught a new sign. They point to it, and they tell us we should know about it. That is more than knowing how it looks. The things out there have become cues, making us aware of the meaning of signs. This means that the reality of humans who have grown up in society is not "the real world out there" as it is perceived through our senses but a reality as it has been constructed by people talking about it. Some philosophers, such as John Searle, prefer to distinguish between a social reality they accept as discursively constructed, and thus as the result of what Searle calls *collaborative speech acts of declaration* (e.g., more or less abstract institutions such as marriage, citizenship, or education), and the natural world, namely, the world to which the natural sciences are said to give us direct access. Social reality consists of facts created, constituted, and maintained through discourse. But the reality

of nature is for Searle not a social construct. He is adamant that things such as hydrogen atoms, tectonic plates, viruses, trees, and galaxies exist independently of people talking about them. This is also the stance of the framework of the sociology of knowledge. In the tradition of Karl Mannheim, the cofounder of this movement, it is claimed that talking about culture, about social institutions, is different from talking about nature. Beliefs and attitudes, law, custom, art, and so on are ideas the members of a social group, a community, have collaborated to construct. They are neither true nor false, and in their place, there could be other ideas. What is more, these concepts tend to change over time. Everything we find in social groups that is not universal (and thus brought about by our genetic heritage) is contingent. However, the sociology of knowledge, and John Searle, claim that what is said about a natural phenomenon is either true or not true.

Other paradigms, for instance, social constructionism, whether in its British variation (e.g., Jonathan Potter) or in its U.S. version (e.g., Kenneth Gergen) deny that dealing with the natural world is systematically different from dealing with the social world. The natural sciences today may employ a methodology that is more elaborate than the classification of so-called primitive societies. But all taxonomies have to be negotiated. They are contingent. We can classify deviant moods in the tradition of Hippocrates as the disturbed homoeostasis of bodily humors or, in today's notions, as the result of hormonal imbalances. The taxonomies we use are outgrowths of discursively negotiated arrangements, not a mirror image of the "real" world. The poststructuralist philosopher Jacques Derrida points out that however far back we go, a language sign, a symbol, never refers to a thing of a discourse-external natural reality but only ever to former discourse events in which this sign has been used. We cannot hope to find out the truth about reality; we can only try to make sense of what has been said.

Discourse and Society From a Constructionist Perspective

So far, we have defined society as people symbolically interacting with each other, and a specific society, or discourse community, as a subset of people who are observed (by themselves or others) to distinguish themselves from other subsets in ways expressed in their specific discourse. But this is only one way to look at society. We can also define it as the entirety of symbolic interactions happening between people, thus excluding individuals from it. For Niklas Luhmann, discourse, or as he calls it "communication," which is, for him, coextensive with society, is a system that is not directly linked to the intentionality of the people contributing to it. We can view society as a system that manifests itself in discourse, a system that in principle does not require to be populated by individuals and their intentionalities. Society in this sense can be compared with a computational system, such as a distributed network of computers doing things together that no single computer could do. All computers together are involved in a process that delivers results according to a program that controls the subprocesses carried out by each computer and coordinates the exchange of data between them. What each individual computer contributes becomes relevant only within the process as a whole. In a similar way, we can define discourse as an emerging system consisting of distributed parts that turn out text after text. It is autopoietic in the sense that that there is no programmer endowed with intentionality who has generated it and assigned meaning to it. The individuals themselves with their intentionalities have no part in this system; they are outside it.

From a radically constructionist perspective, discourses are processes, like apples growing on a tree or computer programs generating haiku after haiku. If computers can do poems, then people and their intentionalities do not have to be part of the system. We can view discourses as running without anyone's agency or intentionality. We can look at a discourse as a random process churning out texts, while at the same time gradually modifying their own rules for what can be said and what cannot. From this perspective, the individuals, the people populating discourse, including the "I" and the "you," are themselves just objects, and all we can know about them (including ourselves) is what is said about them in discourse by the others and by ourselves. While each of us is absolutely certain of our individual unique selfhood (unless we believe ourselves to be schizophrenic), this self-awareness as such cannot be communicated. Of course, people talk about themselves, their feelings, their attitudes, and their opinions. But from the outside, we can never be sure if it is a person endowed with intentionality or a robot imitating such a person. All that is available to us is what we find in discourse, in what this avatar of a speaker says in reaction to what has been said before.

While we normally do not doubt our individual selfhood, everyone else's selfhood is closed to us and inaccessible to us. All we can ever know about the "speaker" is what we are told by him or by others.

Discourse does not necessarily presuppose human intentionalities; it goes on, whether people say they are making sense of it or not. When we are asked by someone how we feel, we normally respond by saying something that has been said before. What for us appears to be the result of our deliberation is, for the hearer, indistinguishable from an arbitrary, random choice from a rather limited list. What is said within the evolving rules constraining what can be said is, from the outside, indistinguishable from randomness. We may well believe that what we say is the result of serious deliberation, but there is no way to verify this claim.

In the perspective of a radical constructionism, intentional agents are not needed for discourse to thrive. The same is the case for society. The framework of philosophical eliminativism, too, assumes that intentionality is a contingent, supervenient feature in the evolution of the mind, indistinguishable from an illusion. As eliminativists see it, society would function just as well without intentionality. Society is a discourse construct, and discourse is an autopoietic system sui generis, a system generating utterance after utterance whether or not there are people consciously deciding to make their contributions. Society is creating sense, in Luhmann's sense, without being in need of someone making sense of it. Society is, for him, what happens between people, and it is only loosely coupled with the people between whom discourse is happening. Perhaps, though, people are not entirely expendable. The innovation on which a healthy discourse thrives can well be seen as brought about by what is the natural (and not the cultural) side of a person. Without the everlasting tensions between nature and culture, discourse might actually come to a halt.

Reconciling Perspectives: Randomness and Intentionality

All contributions to discourse, whether generated by intentional minds or by randomly operating computers, can be analyzed for their meaning. This is called "making sense" of them, or "interpreting" them. People believe that interpretation is something requiring the intentionality of a human mind, actually the intentionalities of all those who take part in the collaborative act of interpretation. For we do not interpret the world of discourse for ourselves; we offer our interpretations normally to the other members of an interpretive community. Making sense is something discourse communities engage in. The outcome of such a collaborative endeavor, however, is unpredictable. Board members, for instance, do not know in advance where the interactions taking place at their meeting will take them. They come equipped with intentions. But soon the "collective mind" of their little community takes over, having them make contributions they had not planned when preparing for the meeting. The meeting will often take a course no one had foreseen. It is as if a sense-making collectivity emerges with its own intentionality.

Making sense of what has been said—of ourselves, others, and the world around us—is not something happening on a higher level of discourse; it is not looking at discourse from the outside. Every contribution to any discourse involves commenting on previous contributions, evaluating them, and viewing them from a slightly different perspective. Indeed, we can look at the diachronic dimension of discourse as adding layer upon layer of interpretation on what was never anything but interpretation in the first place. We take these previous utterances as they are, keen to find out what they mean, ascribing to ourselves, but not to these texts, intentionality, a capability for reasoning, for making decisions based on justifiable intuitions. For their meaning, it is irrelevant if these utterances have been produced by computers or humans. Intentionality is something we only ascribe to ourselves, as members of the interpretive community.

In the same way that we make sense of the diachronic dimension of discourse, we also make sense of social history, indeed of all human history. We can never be sure why an individual or a group of individuals engaged in this or that behavior. We have no way to distinguish what these people did from random acts. But in our endeavor to make sense of them, we apply our intentionality—an intentionality, it has to be repeated, we were not born with but that is the outcome of having been engaged in symbolic interaction. We cannot pretend otherwise. Once we have acquired it, we forego the possibility of viewing what we do as random. Both perspectives, the intentional and the random, are complementary. But they are compatible as well. They view discourse and society not as the necessary results of the workings of inalterable laws. Such laws are our focus when we

attempt to make sense of what happens, for instance, in an ant colony, in a group of chimpanzees, or in the fission of an atom. There we search for common denominators for what we see happening. For discourse and society, our focus is different. What interests us is not what the texts have in common but how they differ. Our focus is the contingencies we observe in each discourse and in each discourse community, regardless of whether these contingencies are due to intentionality or randomness.

Wolfgang Teubert

See also Discourse Analysis; Luhmann's Social Theory; Searle and the Construction of Social Reality; Semantics and Pragmatics; Sociolinguistics; Symbolism

Further Readings

Berger, P., & Luckmann, T. (1966). *The social construction of reality.* Garden City, NY: Anchor Books.

Clifford, J., & Marcus, G. E. (Eds.). (1986). *Writing culture: The poetics and politics of ethnography.* Berkeley: University of California Press.

Fairclough, N. (2001). *Language and power* (2nd ed.). London, England: Longman.

Geertz, C. (1973). *The interpretation of cultures.* New York, NY: Basic Books.

Gergen, K., & Gergen, M. (2010). *Social construction: Entering the dialogue.* Lima, OH: CSS.

Hall, S. (1997). *Representation: Cultural representations and signifying practices.* London, England: Sage.

Lock, A., & Strong, T. (2010). *Social constructionism: Sources and stirrings in theory and practice.* Cambridge, England: Cambridge University Press

Luhmann, N. (2012). *Theory of society* (Vol. 1). Palo Alto, CA: Stanford University Press. (Original work published 1998)

Potter, J. (1996). *Representing reality.* London, England: Sage.

Searle, J. (2010). *Making the social world.* Oxford, England: Oxford University Press.

Teubert, W. (2010). *Meaning, discourse and society.* Cambridge, England: Cambridge University Press.

Language-Games and Forms of Life

This entry introduces the later Wittgenstein's core ideas about language-games and forms of life and reviews in some detail their various aspects. These two notions, along with the related one of rule following (see separate entry on this topic), have been particularly important during the earlier phases of the philosophy of the social sciences in the English-speaking world. This has been especially so following Peter Winch's appropriation of such key notions from the philosophy of Wittgenstein and his use of them in underlining the essential difference between understanding social action and explaining natural phenomena.

Introduction

Ludwig Wittgenstein (1889–1951), among the foremost philosophers of the 20th century, introduced the notion of a language-game. There are three important philosophical uses of the language-game. First, it is a methodological tool in the critical examination of theories of language and mind. Second, it plays a pedagogical role in teaching the child or an initiate learner language and mathematics. Third, it is a descriptive/explanatory device for illuminating the structure of language, in particular the way in which words are bound up with actions.

Language-games are so called because they are analogous to games people play: board games, ball games, gambling games, and so on. Language does not have a single uniform structure, identified by its syntax (grammatical rules for combining words into meaningful whole sentences) and semantics (theory of meaning for words). It is rather a motley of language-games, which overlap, crisscross, and interpenetrate in various ways. Chess and checkers are very similar, but they are quite distant from basketball or hockey. Language is used in lots of different ways, to do different things for different purposes; accordingly, there can be no single theory for language. Like ordinary games, language-games have rules or norms for how one is to play the game; and also like most games, language-games use markers or signs in playing the game.

As a game of chess uses pawns, knights, bishops, and other pieces and as basketball uses a basketball, speakers use words and sentences to make moves within a language-game, to do various things. We use language to get married, describe an itch, recite a poem, or follow instructions. A language-game can be as simple as the child's game of ring-a-ring-a-roses or as complex as algebra or physics. But most interesting are the ordinary language-games through which we live our lives. Games like going to the

grocery store, expressing pain verbally, or expecting someone for coffee. These language-games require, as the necessary background to their being played, our human form of life.

Methodological Role of Language-Games

Language-games are tools for the critical and diagnostic inquiry into traditional philosophical theories of language and mind. There are four primary foci of this scrutiny: (1) names (the way words label objects), (2) rules and rule following, (3) self-knowledge, and (4) consciousness. In each case, Wittgenstein constructs a simple language-game, complete in itself, as a diagnostic tool for identifying the mistakes and incoherence of the philosophical theory under examination. Indeed, he opens his examination of the way in which words relate to objects by introducing his first methodological language-game: the Builders. The Builders is a very primitive form of language-game. Two builders, A and B, use a limited vocabulary (terms for four types of building blocks) to coordinate their behavior. A calls out one of the four terms, and B brings to A the building block corresponding to that term. It is only natural to hold that the four terms mean the four kinds of building stones. This simple language-game is the cornerstone of Wittgenstein's critique of so-called referential theories of meaning.

The methodological role of the Builders (or any other language-game so used) is subject to two important constraints: First, it is an artificially constructed language-game that is designed to fit the requirements of the philosophical theory in question. The Builders game consists of four signs, each of which corresponds to a type of building stone. In short, each word labels a type of building stone. This is one prominent theory of how words relate to the world, but one that is undermined by the use of language-games. The static relation between label and object is belied by the dynamic character of the language-game. The second constraint is that the language-game, selected for methodological use, cannot use proprietary notions in describing the game. If the only way in which a language-game could be described requires technical or proprietary notions, then that language-game simply begs the question in favor of the theory under scrutiny. The language-game provides a simple world in which the explanatory elements of the targeted philosophical theory become transparent to the theorist or critic.

There is a striking similarity between simple language-games and those of the young child just learning language or the pupil just learning mathematics. As the simple language-games are spare and cognitively primitive, so the learning games of the young child or pupil are very simple. The initiate's learning situation—that is, one that involves a learner who is just beginning—is one that helps reveal how rules or norms are acquired and subsequently shape the child's life. Acquiring mastery of language is realized through learning techniques and skills in the *use* of words. This very basic way of learning is closer to a causal process, but it is a causal process that is formed by the child's relation to the adult, one who knows how the skill or technique is to be exercised. Language-games, as techniques for using language, must be learned in such a way that the causal susceptibilities of the child are exploited by the adult in teaching the child rules. The child is trained in the use of words. The learning situation thus bridges the methodological use of language-games and the descriptive/explanatory use. It elucidates the word–object relation within a simple language; and it displays the complex normative and causal interaction that is definitive of language mastery.

Descriptive/Explanatory Use of Language-Games

The third use of language-games is to provide a way of describing our linguistic practices that highlights the ways in which words are integrated into, and so are essential to, our actions. The first thing to note in shifting from the methodological use of language-games to the descriptive/explanatory use is that the notion of a language-game is freed from the conditions of its methodological use. This opens up space to recognize the enormous flexibility of the concept "language-game." It is applicable to an indefinitely large number of cases, which vary in the complexity and kind of action that are involved. The use of this term is not restricted to the simple primitive language-games used in constructing objections to philosophical theories of language.

The second point of contrast with the methodological use of language-games is that it constitutes an alternative picture of language—an alternative, that is, to the standard conception. According to the standard conception, language consists, at base, of a set of formal syntactic and logical rules and a

vocabulary list. Though the specific set of grammatical rules and vocabulary words may differ from one natural language to another, all languages share the logical structure of rules applicable to words in a determinate way. The consequences of imposing this rigid structure on language is that the logical space for all that is sayable is fixed once and for all. On this standard conception of language, differences among the uses of language concern superficial variations, while at a deep level, every legitimate or meaningful use of language is already fixed by the structure of language. By contrast, language as an array of language-games does not have this feature. It is indeterminate as to what will count as a meaningful use of language.

There are a number of ways in which the language-game conception differs from the standard conception. The number of different kinds of language-game is countless. The space for possible new games is open-ended. There is no principled way to draw a boundary around the range or kind of meaningful language-games. Though playing language-games is a normative undertaking—that is, it is rule governed—there is no set of universal rules that governs all uses of language. In emphasizing the multiplicity of games, that speaking a language is part of an activity is highlighted. Language-games, like ordinary games, require the participants to act in certain ways. Once again, we can find an analogy with the child learning simple language-games. The point, in teaching, is not to engage the child intellectually but to get the child to respond in the right ways to certain objects or expressions with his or her own words. Language-games are normatively structured activities, multifarious, showing a *family resemblance* among each other and an *open texture*. The situation of initiate learning is entry into language-games. And entry into a language-game is ipso facto expressive of a form of life, the human form of life.

Form of Life

Though there is no set of rules of language applicable to all language-games, and because the rules of a language-game cannot be fully extracted from the context in which the rules have life, there is no way that rules per se—that is, as abstracted from the activities of the participants—can explain our orientation to the world and to others. Such rules, whether explicitly obeyed or only implicitly followed, exist against the background of our human form of life. This is a much more nebulous notion than that of language-game. It is the idea that there is something like human nature, which is universal to all human beings and indeed to any being who acts and reacts in the ways that we find natural. The human form of life is the locus for what we find similar at the bedrock of each language-game, whether this is the recognition of colors or knowing how to continue the natural number sequence or what behavior to expect from chairs. Our human form of life is the universal context within which we live and agree (or disagree). It is what we share as human beings. It is what is necessary if the novice is to acquire mastery of language. Our shared reactions as to what is similar, our orientation toward our fellow human beings, are exploited in the successful navigation of our various language-games. Our norms, rules, and rule following are embedded in our language-games, in a bedrock of shared judgments of similarity and communal agreement. The human form of life, as the context within which rules can be followed, cannot be successfully abstracted from its active role within language-games.

This foundational role of form of life for language-games reveals the ways in which *trust* and *certainty* are essential to our linguistic practices. In playing chess, we are certain of the board's boundaries remaining fixed. In seeing that another is in pain, we trust that she is not deceiving us. These sound like the banalities that they are. But their implicit role in our linguistic lives, which is to say within our lives, is of fundamental importance. It is not just that language-games are composed of norms and linguistic signs. Playing the games must presuppose certainty with respect to many matters and trust with respect to our social partners. Lose either of these foundational features, and the language-game dissolves. It is this background, the way we human beings are situated in the world, that enables us to become participants in normative practices. When a person sits down in a chair, he expects it to remain solid. When one moves a rook in the context of a game of chess, one is certain that the targeted queen will not fly away. When the student learns history, it remains constant for her that Napoleon was exiled to Elba. Our language-games are anchored in the tacit, and often never articulated, background of certainties that make the moves in the game possible and that reflect the character of our human way of believing.

In sum, the notion of a language-game, especially as enriched by certainty and trust, has an essential connection to our form of life. Our shared form of life requires both a human nature as well as natural facts of the sort that support the language-games within which we live our lives. The certainty expressed in our language-games is not one that has been justified. On the contrary, the certainties of our form of life are like something animal. That is what it is to *be* as of a form of life. The peculiarity of our form of life, as opposed to that of other animals, is that we must live our lives in normatively informed and self-conscious language-games.

Meredith Williams

See also Causes Versus Reasons in Action Explanation; Explanation Versus Understanding; Language, Philosophy of; Relativism and the Social Sciences: From the Sapir-Whorf Hypothesis to Peter Winch; Rule Following; Social Practices; Speech Acts

Further Readings

Baker, G. P., & Hacker, P. M. S. (1980). *Wittgenstein understanding and meaning*. Chicago, IL: University of Chicago Press.

Luntley, M. (2003). *Wittgenstein: Meaning and judgment*. Oxford, England: Blackwell.

McGinn, M. (1997). *Wittgenstein and the philosophical investigations*. London, England: Routledge.

Rhees, R. (1970). *Discussions of Wittgenstein*. New York, NY: Schocken.

Stern, D. G. (2004). *Wittgenstein's philosophical investigations: An introduction*. Cambridge, England: Cambridge University Press.

Winch, P. (2003). *The idea of a social science and its relation to philosophy* (2nd ed.). London, England: Routledge.

Wittgenstein, L. (2009). *Philosophical investigations* (4th ed.; P. M. S. Hacker & J. Schulte, Eds. & Trans.). Oxford, England: Wiley-Blackwell.

Law, Social Phenomenon of

Reflection on the fundamental characteristics of law, and law's relation to morality and to religious edicts, is probably as old as law itself and has been central to Western thought since its earliest recorded history. This entry focuses only on its more recent incarnations. The growth and greater distinctness of legal phenomena, on the one hand, and the rise of scientific anthropology and analytic philosophy, on the other, have given these questions particular urgency in recent times.

The Anthropological Perspective: What Are the Characteristics of a Community Governed by Law?

Probably the first to show interest in identifying the unique features of law were sociologists and anthropologists. Roughly speaking, one can speak of two schools of thought: According to one view, more often voiced by scholars interested in modern societies (e.g., Max Weber), the legal domain is to be identified with formally promulgated rules backed by formal mechanisms of coercion. The other view, typically held by those studying indigenous societies (e.g., the social anthropologist Bronisław Malinowsky), identified law with a set of widely shared norms and explicitly declined to associate law with central authority. When law is understood in this way, it is hard to imagine—indeed, it may be conceptually impossible to conceive of—a social community without law, as arguably what makes a group of people a community is the existence of shared norms. By contrast, understood in the other way, law is a feature of larger and more complex communities but is not necessarily found in every society.

Proponents of the first approach have argued that the latter approach cannot distinguish between law and a host of other norms that govern people's lives; in response, defenders of the latter approach have argued that the focus on formal institutions biases the account in favor of the particular form law has taken in modern states and fails to account for the functionally similar phenomena found in other societies.

In subsequent writings, the divergence in views has roughly followed academic divisions. Among legal sociologists and anthropologists, growing skepticism over "essentialism" and doubts about the point of such inquiries have led to a decline in interest in identifying the distinctive features of law. In its stead has emerged a view that dominates much of social-scientific work—called "legal pluralism." According to a well-known characterization of legal pluralism,

"people experience justice (and injustice) not only (or usually) in forums sponsored by the state but at the primary institutional locations of their activity—home, neighborhood, workplace, business deal and so on" (Galanter, 1981, p. 17). Legal pluralists thus extend the second of the two approaches outlined above to large, modern societies, arguing that even in such societies the normative domain is filled with a multiplicity of overlapping normative structures that are impossible to distinguish; moreover, because of their constant interaction and mutual influence, attempts to isolate one part of that picture will in all likelihood be misleading.

The Analytic Perspective: What Is Law?

In contrast to this inclusive attitude of sociologists, legal philosophers have largely adopted the opposite approach, which identifies law with institutionalized normative structures of formal law. This has been the starting point, more often assumed than explicitly defended, of most of the literature on the "analytic" or "conceptual" question "What is law?"—the question that quickly became the fundamental question of general jurisprudence in modern times. The academic divide has naturally led to a change in the way the question was understood and the methodology adopted for answering it. Instead of examining indigenous societies or attempting to identify historical trends, the favored approach has been philosophical reflection and conceptual analysis.

In the English-speaking world, by far the most influential book in this field remains H. L. A. Hart's *The Concept of Law,* which revived (with significant modifications) the analytical jurisprudence of John Austin. Hart famously offered a quasi-historical account of the emergence of law, according to which the move from prelegal to legal society takes place when society adds to its list of prohibitions and prescriptions (which he called "primary rules") a set of "rules about rules" ("secondary rules") concerned with identifying, changing, and adjudicating disputes about primary rules.

Hart's narrative was not meant to describe historical reality; it is better understood as a thought experiment that ties the existence of law to the existence of certain institutions and formalized characteristics: Law becomes a reality when one finds certain institutions that turn an amalgamation of norms (a "set" of rules in Hart's language) into a "system."

Despite being a paradigmatic study of the approach that sought to distinguish law from other normative structures, Hart's account could be embraced by some legal sociologists who favored the alternative view. Because his account focused on *rules* rather than *institutions,* such thinkers could adopt Hart's view by arguing that there were societies governed exclusively by primary rules. Even in societies that lack clearly defined formal legal institutions (legislature, judiciary, etc.), one could identify rules dealing with change, dispute resolution, and so on.

This convergence in view and cross-disciplinary attention is perhaps not surprising. In Hart's work, one finds fleeting reference to the sociological work, and when he was writing, the disciplinary boundaries were less clearly defined than they are today. In subsequent philosophical work, the divide between the sociological and philosophical work has become more pronounced. The boundaries manifested themselves not just in a growing methodological divide but also in a subtle shift in the question asked: The sociological concern with describing the distinct features of a *society* ruled by law (as opposed to one without law) has become the more distinctly philosophical interest in identifying the distinct features of *law itself*. Even though this underlying question has sometimes become obscured in some highly technical and abstruse debates, arguably this is what has been at the heart of the debates on questions such as the relationship between law and morality, the place of coercion and sanctions in law, and the difference between a legal regime and that of a crime syndicate, all issues that preoccupied legal philosophers since the publication of Hart's book.

It is in this context that the search for the "nature of law" as the search for the existence conditions of law was born, in particular the debate that has been at the heart of philosophical thinking about law in the past 50 years: the debate between legal positivism (LP) and antipositivism. (LP is often contrasted with natural law, but this contrast is historically inaccurate and in the present context somewhat underinclusive if it is to include some of the most prominent antipositivists.) In Hart's work and in that of many others, LP was meant as a rather general rejection of a necessary connection between law and morality, but in the context of the question "What is law?" it would be useful to distinguish between four different senses, in which one could reject a necessary

connection between law and morality. The first two are methodological: One is the view that legal philosophy is distinct from political and moral philosophy, as it is concerned with providing a "description" of the nature of law. More precisely, it is sometimes stated that one need not rely on any other evaluative considerations (moral or otherwise) when providing a theory of law. In the second methodological thesis, LP is used with a meaning close to the way "positivism" is used by sociologists following Auguste Comte—that is, as the thesis that the methods of the natural sciences should be extended to the social sciences. Hart and many of his followers adopted the first methodological thesis but rejected the second. In this, incidentally, they stand in sharp contrast to Thomas Hobbes and Jeremy Bentham, sometimes considered the founders of LP, who rejected the first and adopted the second, and to most contemporary antipositivists, who rejected both methodological theses (Brian Leiter may be an example of someone who accepted both).

In addition to these methodological theses, we can identify at least two additional substantive LP theses. One is that the existence conditions of legal norms, the terms of legal validity, do not necessarily contain moral conditions (with other proponents of "hard" or "exclusive" LP taking the stronger stance that the terms of legal validity necessarily do not contain moral conditions). The second substantive LP thesis in Hart's work is that law's normative force, the sense in which it creates obligations, does not depend on moral premises. Again, one finds different views among legal philosophers, with some (e.g., Hart) adopting both, some (e.g., Joseph Raz) adopting only the first, and antipositivists like Ronald Dworkin rejecting both.

While these questions preoccupied the literature for the first few decades after the publication of Hart's main work, more recently, and especially after the publication of the second edition of the book, there has been growing interest in the methodological LP theses. Here, the lines are drawn somewhat differently, but generally speaking, there are important links between certain methodological and substantive positions. Thus, the alternative, antipositivist perspective is best understood not simply as a rejection of certain substantive views about what law is but more fundamentally as an alternative view about how one should go about answering this question. On this view, one cannot explain the nature of law outside a certain moral or political context. What makes something into law does not depend on a certain institutional framework but must be explained as part of an account of what makes for a legitimate political order. Some have sought to show that law must be understood as part of a substantive theory of the good (e.g., the Oxford legal philosopher John Finnis) or a legitimate political order (e.g., Dworkin), while others have laid more emphasis on the procedural aspects of law (as Lon Fuller does, and Tom Tyler, from a sociological perspective). These thinkers all reject the view that legal philosophy can be descriptive.

These differences notwithstanding, the prevailing view among legal philosophers on both sides of the positivist/antipositivist divide sees legal philosophy as concerned with fundamentally different questions from those of social scientists. This attitude has led to a neglect of potential connections, more explicitly pursued in the past, between legal philosophy and social-scientific study of law.

The Institutional Perspective: The Limits of What the Law Can Do

Alongside these conceptual inquiries on the boundaries of law, there emerged a more practical concern with the limits of legal action. (This question is distinct from, although possibly connected to, the *moral* question of the limits of law, namely, whether there are certain things law should not be used for, even if it can be used effectively against them.) This question presupposes the multiplicity of norms governing society, as emphasized by legal pluralists, but contrary to them, it assumes the possibility of distinguishing, however imprecisely, between legal and nonlegal norms. Given the fact of normative plurality, the question such scholars ask is the extent to which law can be used together with, or in opposition to, nonlegal norms.

As a positive matter, various scholars (e.g., Robert Ellickson and Stewart Macaulay) have criticized what they perceive as lawyers' excessive focus on law and the distorted image they paint of its impact. Such works point out that in many contexts legal norms are often ignored in favor of nonlegal norms with a different content. From a more critical perspective, work on the interaction of social norms and the law has often cast doubt on the ability of legal norms to bring about significant changes to

prevailing social norms (as Gerald Rosenberg has explored). Law, it is argued, is quite often a follower of nonlegal norms rather than their generator, and when its edicts differ markedly from existing social norms, they are likely to be ignored. These controversial claims are difficult to assess in general, and it is possible that the seemingly more "technical" legal rules that often receive less attention (e.g., tax or welfare provisions) have greater power to change social norms in less direct ways.

Be that as it may, such claims may seem quite distinct from the questions philosophers are interested in, but there are important connections between them. At one level, such claims may vindicate some forms of LP, for example, the claim that it is possible to identify the "limits" of law (as Raz has argued). In a different sense, however, such claims may support more antipositivist views of law as they show that law cannot depart significantly from social norms and that the content of legal norms is not fixed by legal materials alone. This in turn suggests a stronger link than LP assumes exists between what counts as law and the content of legal norms, between the content of legal norms and social acceptability, and between the latter and questions of legitimacy.

Conclusion

Perhaps all this suggests that in identifying and understanding law we must acknowledge the truth of seemingly conflicting perspectives. Positivists argue that what distinguishes law from nonlaw are certain structural features; this leads them to a perception of law as a morally neutral social phenomenon and correspondingly conceives of legal philosophy as broadly separate from moral and political philosophy. By contrast, on antipositivist views, what distinguishes law from nonlaw is the fact that legal edicts, unlike nonlegal ones, are potentially legitimate. Unsurprisingly, this approach blurs the dividing line between legal, moral, and political philosophy. Within the confines of philosophical inquiry, these two perspectives seem to be self-validating, hence the mutual incomprehension between the positivist and antipositivist camps in jurisprudence. The key to breaking this impasse may be in recognizing that law is both the product of an autonomous domain, with its set of technical rules that allow lawyers from different jurisdictions to converse more easily with one another than with nonlawyers within their community, and a product of a particular culture bearing the distinct marks of the community from which it emerged. Likewise, it has both institutional, structural features that may be relatively universal andnormative, justificatory ones that are likely to be more local. Greater recognition of the complexity of legal phenomena and openness to the links between philosophical and social-scientific work on law could prove helpful in addressing these questions.

Dan Priel

See also Laws Versus Teleology; Legal Epistemology; Social Institutions; Social Norms; Social Rules

Further Readings

Austin, J. (1995). *The province of jurisprudence determined* (W. E. Rumble, Ed.). Cambridge, England: Cambridge University Press. (Original work published 1832)

Black, D. (1973). *The behavior of law.* New York, NY: Academic Press.

Dworkin, R. (1986). *Law's empire.* Cambridge, MA: Harvard University Press.

Dworkin, R. (2006). *Justice in robes.* Cambridge, MA: Harvard University Press.

Ellickson, R. C. (1991). *Order without law: How neighbors settle disputes.* Cambridge, MA: Harvard University Press.

Finnis, J. (2011). *Natural law and natural rights* (2nd ed.). Oxford, England: Oxford University Press.

Fuller, L. L. (1969). *The morality of law* (Rev. ed.). New Haven, CT: Yale University Press.

Galanter, M. (1981). Justice in many rooms: Courts, private ordering, and indigenous law. *Journal of Legal Pluralism, 19,* 1–47.

Griffiths, J. (1986). What is legal pluralism? *Journal of Legal Pluralism, 24,* 1–55.

Hart, H. L. A. (1994). *The concept of law* (2nd ed.). Oxford, England: Clarendon Press. (First edition published 1961)

Leiter, B. (2007). *Naturalizing jurisprudence: Essays on American legal realism and naturalism in legal philosophy.* Oxford, England: Oxford University Press.

Macaulay, S. (1965). Non-contractual relations in business: A preliminary study. *American Sociological Review, 28,* 55–67.

Raz, J. (1972). Legal principles and the limits of law. *Yale Law Journal, 81,* 823–854.

Raz, J. (1995). *Ethics in the public domain: Essays on the morality of law and politics* (Rev. ed.). Oxford, England: Clarendon Press.

Raz, J. (2009). *The authority of law: Essays on law and morality* (2nd ed.). Oxford, England: Oxford University Press.

Rosenberg, G. N. (2008). *The hollow hope: Can courts bring about social change* (2nd ed.). Chicago, IL: University of Chicago Press.

Tamanaha, B. Z. (1995). An analytical map of social scientific approaches to the concept of law. *Oxford Journal of Legal Studies, 15,* 501–535.

Tyler, T. (1990). *Why people obey the law.* New Haven, CT: Yale University Press.

Laws of Nature

After centuries of neglect, an intense interest in scientific laws was generated in the mid 20th century, principally by the seminal work on scientific explanation by the philosopher of science Carl Hempel. It became evident that the concept of law had an important role not only in the philosophy of science but more generally for the philosophy of mind, psychology, and the social sciences.

The entry gives an overview of two ways of explicating what a "law of nature" is and draws connections with the social sciences.

Definitional Approaches

The now numerous accounts of scientific law fall into two distinct types. One type we might call *definitional* because it involves a familiar way of clarifying a concept—by giving a definition. That involves filling in the blank in the expression

"A is a law of nature if and only if . . . ,"

so that the resulting sentence is one that must be true. There are at least four known proposals of this definitional type:

1. A is a law if and only if it is a general empirical conditional.

A conditional is an "if . . . then" set of sentences. A is a law (say "All metals expand upon heating" or "If a metal is heated, then it expands") for which there is evidence of a strong kind. This is usually called the *Humean account of laws,* after the 18th-century Scottish philosopher David Hume. It has been rejected on the grounds that a law not only expresses what happens to be the case, but it also expresses what must be the case, where this "must" is a special requirement of physical necessity.

2. A is a law if and only if A is a general empirical conditional and it is physically necessary.

There have been many different philosophical accounts of physical necessity. Sometimes those accounts begin with the intuitive idea that if it is a law that all metals expand when heated, then not only is it false that there are metals that do not as a matter of fact expand upon heating, it is also not physically possible that there are such cases. Sometimes this concept is explained by requiring that something is physically possible, if it is not ruled out by laws. But this explanation relies on the concept of law and cannot be used to define the notion of scientific law, on pain of circularity. Consequently, defining the modal notion of physical possibility in contrast to metaphysical and logical possibility is an open and important problem even today.

There is a view of laws that marks a radical contrast with the preceding accounts. This is the Armstrong–Dretske–Tooley view, according to which (with slight differences between the three accounts), laws are not conditionals. That is,

3. A is a law if and only if it is a special kind of "necessitation" relation between universals.

There are several versions of this account. On one version, we are told that one special kind of property (a universal) generates or brings about another, but there is no hint of what is meant by "generation." On another, the relation is supposed to be one of causation that holds between two states of affairs. It is difficult to see how any of the standard laws fit this pattern. In fact, on the first version, the universals or properties that enter into laws are required to be true of things that exist. The result is that Newton's Law of Inertia ("If there are no forces acting on a body, then that body will not accelerate") fails to be a law. It is serious when the grandfather of all laws doesn't satisfy the condition for being a law.

The next account ("the best-systems account") is probably the most widely accepted account at present. It originated with the 19th-century English philosopher John Stuart Mill and, with subsequent refinements, is now known as the Mill–Ramsey–Lewis account. Consider any collection of true sentences. There will be many such collections. Some

will be simpler than others, and some will contain more information than others. Choose the one that is the best combination of simplicity and information. The best-systems account then defines the concept of a law in the following way:

4. If A is any empirical generalization, then A is a law if and only if it is a member of the best collection of true sentences.

If there are several collections that are the best combination of simplicity and informativeness, then this account is modified so that "A is a law if and only if it belongs to all of them."

There are several problems. There is the issue of how to understand the required notions of simplicity and informativeness that are not specified. If different people specify what is meant, then the concept of law will vary with those differences. That seems to speak against the objectivity of the best-systems account. Another difficulty is that we do not know at any given time whether a particular empirical generalization is a law. To determine that, one has to know whether that generalization belongs to the best system of truths, and at any given time, we do not know what that system is. We may have a hunch, we may guess, but we do not know.

A further problem with this account motivates a fifth account. There is the possibility that various different sciences may have laws of their own, but the requirements for being a law of physics, psychology, biology, or economics may be different. The best-systems account does not allow for the possibility that the concept of scientific law may be different for different sciences. This possibility may be important for those scholars who work in any of the social sciences for whom the standard for being a law of physics, for example, may not be appropriate for social sciences such as sociology, social psychology, anthropology, political science, and economics. Thus, there is another definitional type of account:

5. If A is any empirical generalization and T is any theory, then A is a law of the theory T if and only if A plays a distinctive special role in that theory.

There are several ways in which the "distinctive role" has been understood. There is the early attempt by Richard Braithwaite, in which the special role is a deductive one within a deductive system that includes the theory. A more recent variation of this theme is that of John Roberts, who explains the special theoretical role of laws as follows: Associated with any theory there are certain physical quantities, and a general empirical statement is a law of a theory if an only if it is a logical consequence of certain reliability conditions of measurement.

Conceptual Approaches

We turn now from the definitional approach but consider instead those attempts to clarify the concept of law by relating it to other concepts and uniting the whole into a general framework—a minitheory of sorts. The result is not meant to be a definition. We shall call this approach *conceptual*. To illustrate what is meant, we shall list several conditions that, taken together, state how the concept of law is related to those other seminal concepts. It is not precluded that still other ways could be added to the theory to make it less "mini."

1. Laws are true general statements that are empirical.

This requirement is widely adopted, and for those laws that occur in explanations, truth is a consequence of the requirement that anything occurring in an explanation has to be true.

2. If A is a law, then its corresponding counterfactual must be true.

This too is also widely regarded as a hallmark of laws. It holds for laws that are conditional in form, such as the example we saw above: If anything is a metal, then it expands when heated. A counterfactual is an assertion contrary to actual fact—that is, to what actually is the case. So the corresponding counterfactual in this case is the conditional: If anything were a metal, then it would expand upon heating. This condition reflects the view that the first conditional seems to be about those things that in fact are metals, while the counterfactual expands that claim to the wider case, where not only all things that are in fact metals but also anything that could possibly be a metal would expand upon heating. It has to be said that some might not find this motivation convincing but would endorse the condition anyway.

3. If A is a law, then it is not an accidental generalization.

This condition is also regarded as a hallmark of laws (i.e., that unlike an accidental generalization, a law supports a counterfactual conditional), though it has to be added that, to date, no one has given an adequate account of the concept of an accidental generalization. Nevertheless, the usual support for this requirement relies heavily on intuitive examples. Thus, it is sometimes argued that the statement "All the people in a particular room are intelligent," even if true, is not a law because it is clearly a case of an accidental generalization. Well, maybe so. It may seem obvious that it is. Nevertheless, there is still the question as to what considerations show that it is accidental.

The next condition will be familiar from our discussion of it under the definitional approach. Nevertheless, it ought to be listed under the conceptual approach as a consideration of the connection between being a law and being a necessary generalization. That is,

4. If A is a law, then it is also necessary.

Last, we come to a condition that relates the concept of *explanation* with that of laws. It is a somewhat controversial matter whether all explanations involve laws. This may be a problem for some of the social sciences. If it was true that some social sciences have no laws to speak of, then it would follow that those sciences have no explanations to speak of. And that would be a serious matter. We do not wish to decide that matter here. However, there is a condition that, if endorsed, would indicate one very important use of explanations. The condition is this:

5. If A is an empirical generalization and there is some explanation of A, then A is also a law.

In short, even if explanations may not always use laws, they would always yield laws whenever they explain empirical generalizations.

This condition may have important consequence for the social sciences. If it is endorsed, and if it is assumed that although the social sciences have empirical generalizations, none of them are laws, then it follows that the social sciences do not have any explanations. That contentious conclusion seems to be unwarranted. A fuller discussion is beyond the scope of the present entry, but it is worth noting that although philosophers such as Sir Isaiah Berlin have argued that the social sciences do not have any laws, other philosophers, such as Sidney Morgenbesser, have argued forcefully that there are what he called "causal schemata" in the social sciences as well as in the physical sciences, which are very intimately connected with laws. Although they are not themselves laws, they are close cousins. Moreover, Morgenbesser has argued that some social sciences have ideal-type explanations, and in his account of them, he has also argued that they sometimes appeal to general ideal-type laws about human behavior.

Arnold Koslow

See also Causal Explanation, in Philosophy of Science; Causation, Philosophical Views of; Causation in the Social Sciences; Explanation, Theories of; Laws Versus Teleology; Scientific Method

Further Readings

Armstrong, D. (1983). *What is a law of nature?* Cambridge, England: Cambridge University Press.

Berlin, I. (1954). *Historical inevitability* (p. 74). London, England: Oxford University Press.

Braithwaite, R. (1954). *Scientific explanation.* Cambridge, England: Cambridge University Press.

Carroll, J. (Ed.). (2004). *Readings on laws of nature.* Pittsburgh, PA: Pittsburgh University Press.

Dretske, F. (1977). Laws of nature. *Philosophy of Science, 44,* 248–268.

Hempel, C., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15,* 135–175.

Kolsow, A. (2004). Laws and possibilities. *Philosophy of Science, 71,* 719–729.

Martin, M., & McIntyre, L. C. (Eds.). (1994). *Readings in the philosophy of social science.* Cambridge: MIT Press.

Morgenbesser, S. (1956). *Theories and schemata in the social sciences.* Ann Arbor, MI: University Microfilms.

Morgenbesser, S. (1968). Scientific explanation. In D. L. Sills (Ed.), *International encyclopedia of the social* sciences (Vol. 14, pp. 117–122). New York, NY: Macmillan.

Roberts, J. (2008). *The law-governed universe.* New York, NY: Oxford University Press.

Laws Versus Teleology

The word *teleology* derives from the Greek *telos* (τέλος) meaning "end" (goal, aim, completion, fulfillment, or perfection). *Telos* includes the best

condition, activity, or operation of a natural or human agent, such as a prize-winning animal or an excellent (virtuoso) performance by a performing artist. The telos of a process is that for the sake of which each stage occurs. It is the final and most authoritative cause. Most generally, teleology is the account (*logos*) of the ends at which human beings, nature, history, and God aim.

This entry is restricted to two of these: (1) teleology in human action and (2) teleology in nonhuman nature. Its telos is to explain the modern (post–17th-century) rejection of teleology in favor of nonteleological laws—both natural (laws of physics) and human (legislative acts). Even so, this is a large terrain, and the constraints of an entry require considerable truncation of the subject matter. However, an important core theme in the philosophy of social science has been the issue of part–whole relations (as, e.g., in the form of holism versus individualism), and this is an issue that is also central to teleology. Similarly, the idea of mechanical causation or causal mechanisms in general plays an important role in social-scientific explanation. Therefore, the discussion of the contrast between mechanical laws and teleology, though it may be misleadingly thought to be of importance only to the emergence of modern science, is also illuminating in the philosophy of social sciences, where part–whole relations as well as social mechanisms and causation are central. The discussion can also shed light on the issue of whether social institutions, groups, or organizations can be properly said to exhibit virtuous behavior.

Background: Natural Form and Human Virtue

The rejection of teleology in modern philosophy and science can best be understood against the background of the premodern, teleological understanding of nature and human law. Emblematic of the teleological understanding of nature is the notorious Aristotelian term *form*. The form of a natural substance, for example, of a cat, "is [its] nature more than the matter"; form (cat-form) is the "end (telos) and that-for-the-sake-of-which" in the generation of the cat (Aristotle, 1995, 193b7; 1999, 1015a12). Emblematic of the teleological understanding of human law is the term *virtue* (Greek *arête:* αρετή, hence "virtuoso"), for "virtue [e.g., moderation] must be a care for every city" (Aristotle, 1984, 1280b7); indeed, "the proper effect of law is to lead its subjects to their proper virtue" (Aquinas, 1948, I-II, q. 92, a. 1, resp.). We must make an attempt, however brief, to understand these obsolete doctrines.

Natural Form

Form and matter are the constitutive principles of natural substances, which are noteworthy because, unlike artifacts, natural substances "all appear to have *within themselves* [italics added] a principle of motion and rest. . . . So nature is a cause and principle of motion and rest *in* [italics added] that to which it belongs, primarily and essentially and not accidentally" (Aristotle, 1995, 192b14–15, 192b21–23). The complicated adverbial phrase "primarily and essentially and not accidentally" means (among other things) that natural form is internal to the moved thing in a way that cannot be fully derived from, or completely reduced to, its material parts: "[Natural] things will be neither without matter nor determined by their matter"(Aristotle, 1995, 194a14–15). Form is thus a holistic principle. A watch, in contrast, lacks the substantial unity of a cat; it can be taken apart, its parts can be analyzed in isolation from each other, and then both the parts and our knowledge of the parts can be re-aggregated in the whole or even rearranged in a new and different whole. The clock is thus a paradigm for the reductionist or mechanical and anti-Aristotelian conception of the whole–part relation.

The Aristotelian conception of the whole–part relation in natural compounds implies limits to the alteration, manipulation, and transformation of bodies by separation and recombination of their parts. Within this understanding, form is not only a source but also a limit of our knowledge and control of nature (Aristotle, 1999, 1022a5–10)—something that will get rejected by modern scientific thought, as we shall see below, toward the end of the entry, with repercussions for human form and its possible (nonlimited) manipulation (as a river flowing) too.

Form specifies a visible natural kind, or *species,* especially a living species, in which case form is also called *soul*. Form is understood to be the source of the behavior that is most distinctive and characteristic of the species—for example, the distinctive, feline way of moving and meowing specific to cats. Consider the specific difference that is perhaps of greatest importance—the use of words by us human animals and the immense impact of their meanings on the

way we behave; in this, we are unlike any other species. The salient feature determined by the form of a species is also the good—in its exercise of the telos—of that species: "For what is proper to each kind of thing is best and most pleasant for it" (Aristotle, 2002, 1178a5). So it is good and pleasant for the cat to act in that feline (not canine) way, and it is best (and most pleasant) for us humans to use words well for the sake of truth in thought and action. Form is thus a principle that is *species specific,* a term that will acquire a sharper sense when contrasted, below, with the *species-neutral* laws of physics. In any case, form can be at work only in appropriate matter; "for a different form, a different matter, [and so] for each motion it is the subject capable of that motion which has that motion" (Aristotle, 1995, 194b9, 251a14). Now, of all the natural motions visible to us human beings, it is the ones in the heavens, the celestial motions, that are the most distinct, for they are orderly and (as far as we can see) never cease, unlike all earthly changes. But eternal motion can take place only in appropriate—and thus eternal—matter, the indestructible matter of the crystalline spheres in which the visible lights in the heavens must be embedded. Thus, the Aristotelian and Scholastic doctrine of natural form—suited well enough to the heterogeneity of biological phenomena—also concludes to the heterogeneity of terrestrial and celestial matter. And so we can foresee a major reason for the rejection of the teleological account of nonhuman nature: The modern Scientific Revolution, starting in the 17th century, specifically the success of Isaac Newton's theory of universal gravitation, entails the essential homogeneity of celestial and terrestrial matter. Let us sum up the key result on the meaning of natural form in premodern philosophy.

Things that move, change, and behave in essentially different ways as manifested to our senses possess correspondingly different materials and sources of motion. They have different natures. This means that what the different natural kinds have in common, such as mass, is not as fundamental as what differentiates and specifies them. The opposite notion—that what the *apparently* different kinds (e.g., celestial and terrestrial, human and nonhuman) have in common *is* more important than what differentiates and specifies them—is a defining characteristic of *modern* natural science.

Against the premodern background, the significance of Darwinism stands out in stark relief: The successes of evolutionary biology show that the forms (species) are not causes and principles but effects and products (of variation and selection) and, in the current interpretation of the theory, that they are so not according to any purposive (teleological) intention of nature or God but merely by accident.

Human Form and Human Law

It remains to discuss the premodern, teleological understanding of human law: What can it mean to claim that "the proper effect of law is to lead its subjects to their proper virtue"? After all, we just said that the form of any species determines its characteristic behavior (e.g., from cat-soul flows feline behavior, from dog-soul, canine behavior). Therefore—by this analogy to other animal species—human-soul should be the sufficient source of our behavior. But the analogy to other animal species fails; we humans do not operate simply from natural instinct. Rather, the human form is open to a great variety of behaviors according to our very different *dispositions* or character types. Aristotle makes this clear in *Nicomachean Ethics* (2002, 1176a8–12).

In this premodern, Aristotelian account, a person's disposition or character has four sources: (1) individual traits from birth, (2) childhood training, (3) the laws and customs of the community, and (4) one's own past choices. Of these, the third source gives rise to the teleological understanding of human law—that is, law as *formative* of character, of the stable capacity for making good or virtuous choices and avoiding bad or vicious ones.

Character refers to the way we come over time to have our emotions and appetites, thus the way we come to be *disposed* toward and act in the face of the pleasures and pains to which we are subject as animate, rational (possessing speech) individuals in a political community. Our dispositions in turn give rise to our characteristic patterns of choices. Disposition is thus both effect and cause of choice. We make choices for the attainment of ends or goods whose appearance to us is a function of our disposition. (Alcoholism and the alcoholic disposition illustrate the sense of the preceding sentences.) To choose well, therefore, we must be rightly disposed toward pleasures and pains in order to be free of extremes of passion that distort perception and judgment. We must be disposed to the *virtuous mean* between excess and deficiency of passion such that action is

correspondingly appropriate to the particular situation. For example, how do people come, over the course of their lives from childhood to adulthood, to experience fear in facing danger? If slight dangers cause great fear and result in flight, one has acquired a cowardly disposition. If great dangers and risk cause little or no fear, leading to foolhardy actions, one has acquired a rash disposition. The disposition lying in the mean between cowardice and rashness is the virtue of courage. To the defective extremes correspond the distorted perceptions: "The courageous person appears rash to the coward, and cowardly to the rash" (Aristotle, 2002, 1108b20). The courageous person feels, perceives, and acts in the mean "for the sake of the noble, for this is the end (telos) of virtue" (Aristotle, 2002, 1115b12–13). Laws mandating and directing military training lead soldiers to their proper virtue, namely, courage and combat effectiveness. For the problem of the noble, see below.

It is important to emphasize that not only individuals but also groups have dispositions; thus, we speak of the *culture* or *mentality* of a corporation or a country. A person's or group's character is the set of dispositions (e.g., toward dangers, bodily pleasures, work, leisure, money, power, honor, perceived slights, one's own worth, the fortunes of others, humor) that they have acquired over their individual or collective history.

As we depart from this ancient teaching on law, consider whether there is not significant truth to it. But—as Aristotle (2002) himself makes clear—there is also a serious problem: "The courageous person acts for the sake of the noble" (1115b12–13), yet "the noble and the just, about which politics investigates, involve great disagreement and irregularity" (1094b14–16). We can thus easily imagine courageous men getting themselves and others killed over clashing conceptions of the noble, the just, and, let us add, the holy—as happened in the religious wars of the 16th and 17th centuries and as happens again today. The transition from the understanding of law as formative to the nonteleological (and secular) doctrine of law as *limitative*—that is, that law should merely limit the actions of others against my life, liberty, and property—has much to do with this unfortunate history of violent conflict. Concurrent with the religious wars, however, is another salient historical development, the Scientific Revolution, which promised vastly enhanced human power over nature. Essential to the new science is a new type of intelligible principle: the mathematical and experimental law of nature. The modern rejection of teleology in favor of nonteleological natural and human laws may thus be seen to derive in significant part from the desires for civil peace and for technological power.

Laws of Nature Versus Natural Forms

Isaac Newton's law of gravitational force (Newton, 1969, Book 3, Proposition 7 and Corollary 2) is written today as $F = -GMm/R^2$, where M and m are the masses of any two bodies, R is the distance between them, G is the universal gravitational constant, and the minus sign indicates that the force, F, is attractive. It is a paradigm for the laws of classical physics and is remarkable for the new type of universality that it exhibits. Normally, the way two bodies interact depends on what kind or species each is. A dog and a cat, for example, interact in a certain way, a cat and a mouse in another. In fundamental contrast, the law of gravity expresses a principle of motion in nature that is independent of the kind, size, shape, internal structure, and function—in other words, the nature—of the two interacting bodies. For *all* bodies—celestial and terrestrial, natural and artificial, living and nonliving, human and nonhuman—possess mass and relative position, and also velocity, momentum, acceleration, kinetic energy, and so on. These measurable and mathematically representable terms of classical physics are common to all bodies and their parts, unlike Aristotelian forms. As such, they are form or species neutral. Principles of motion that can be expressed in such terms—the mathematical laws of physics expressed in equations—are thus species neutral. The modern philosopher Baruch Spinoza (1985) provides a most succinct formulation of species neutrality in its opposition to natural form:

> That which is common to all . . . and which is equally in a part and in the whole [e.g., mass] does not constitute the essence [i.e., the species in Aristotle's teleological terminology] of any particular thing. . . . Those things which are common to all . . . cannot be conceived except adequately. (II.37–38)

In contrast, forms cannot be conceived except confusedly, due to Aristotle's overreliance on unassisted

sense perception. Thus, to the modern mind, an Aristotelian form must be rejected.

Although it is not expressed in terms of mathematical laws, Darwinian biology is species neutral in the above sense—that is, what is common to all living bodies are the principles of random variation and natural selection. The specific (i.e., in the sense of *species*) behaviors and interactions of the cat, dog, and mouse are thus understood to be ultimately reducible to the universal drive for reproductive success. Similarly so for human behavior as conceived by evolutionary psychology. It is because of classical (atomic-mechanical) natural science, such as particle physics, and Darwinian (gene) biology that species neutrality is a defining characteristic of modern natural science. This means that no account of the human on its own species-specific terms, for example, in terms of speech about the noble, the just, and the holy, can count as scientific today.

A second remarkable feature of Newton's law of gravity is that it (together with his second law, $F = ma$, (where m is mass, a is acceleration, and F is force) enables us to calculate, and thus predict, the trajectory of a body moved under gravitational force if we know or deliberately produce the body's position and velocity at the initiation of motion—for example, rockets. Humanly controlled space flight—impossible according to ancient and medieval astronomy—becomes a reality through mathematical laws of physics. Here, we see the relation between a law of nature and mastery of nature, a major theme in the works of one of the principal fathers of modern science, Francis Bacon, especially in his *The New Organon* (1620/1960). There we find a severe critique of the premodern understanding of nature:

> When man contemplates nature working freely, he meets with different species of things, of animals, of plants, of minerals; whence he readily passes into the opinion that there are in nature certain primary forms which nature endeavors to educe. . . . [But in] nature nothing really exists besides individual bodies [true particles], performing pure individual acts according to law. (I.66, II.2, II.8)

Baconian laws go hand in hand with material reductionism, the idea that sensible wholes are no more than collections of subsensible particles that interact according to the laws. This directly rejects Aristotle's holism of natural form, described above. Indeed,

> forms are figments of the human mind, unless you call those laws of action forms. . . . [Now] whosoever is acquainted with forms [i.e., laws] embraces the unity of nature in materials the most unlike, and is able therefore to detect and bring to light things never yet done, and such as neither the vicissitudes of nature, nor industry in experimenting, nor accident itself, would ever have brought into act. (Bacon, 1620/1960, I.51, II.3)

Bacon's audacious reasoning is that if the true intelligible principles are species neutral, then the apparent heterogeneity of species (evident to our senses) is not rooted in the essential nature of things and, furthermore, might not be a barrier to our operation. For the founders of modern science, the point is not to comprehend the world but to transform it, "whence there cannot but follow an improvement in man's estate" (1620/1960, II.51).

A major anti-Aristotelian, antiteleological premodern voice had been medieval nominalism, which emphasized the radical contingency of all things before the immense power of the biblical creator God and, thereby, undermined the Aristotelian notion of natural form as a stable, intelligible principle. But it was still a long way from the contingency of nature in relation to God's power to the malleability of nature in the face of human power.

Human Law as Limitative Not Formative

The core of the rejection of the teleological account of human law is that it is based on a false understanding of human behavior: If we are honest with ourselves, we will see that our motives are powerfully and irreducibly self-centered. It is unrealistic—and will be ineffective or counterproductive—to think that weak reason can control or moderate the selfish passions. The most fundamental human passion is the compulsory drive for self-preservation, after which come the desires for comforts (convenience), the pleasures, and wealth. These drives cannot be gainsaid; they will move human beings necessarily like the physical forces that move a river, whose flow is involuntary (as Nicoló Machiavelli describes it in *The Prince*, 1985, chap. 25). But human behavior, like the flow of the river, can be directed through

the clever imposition of channels that bring beneficial effects (like cultivation) and avoid harmful ones (like flooding) (compare the contrasting case above in the first section). In the case of human behavior, the channels are well-enforced laws that aim not at "proper virtue," such as courage and moderation, as in the old Aristotelian account, but at channeling the selfish drives of individuals to good public effect (peace and prosperity). Thus, for example, "private vice is public virtue," through legally protected enterprise accompanied by tax laws with teeth in them—that is, in a quasi-mechanical way. As John Locke (1952) puts it memorably, "That prince who shall be so wise and godlike as by established laws of liberty to secure the honest industry of mankind, against the oppression of power and narrowness of party, will quickly be too hard for his neighbors" (sec. 42).

Most generally, individuals are to be guaranteed a protected space in which to pursue happiness as each understands it, consistent with an equal right of others to do the same. Law becomes like a fence around my property, limiting the action of others against me and of me against others without our mutual consent. In this contractual understanding, law "is the direction of a free and intelligent agent to his proper interest" (Locke, 1952, secs. 17 and 57). There is no concern here with disposition or character, for the necessary (involuntary) forces of human self-interest are virtue neutral, somewhat like the forces of nature that are species neutral: Different visible patterns of both natural and human behavior arise not from different internal causes (celestial and terrestrial natures, courageous and cowardly dispositions) but from one and the same forces (gravitation, self-preservation), specified only by differing external conditions. Necessitarianism is thus common to both nonteleological accounts: the natural scientific and the human.

Despite its success in the protection of individual liberty and the production of wealth, the bleak realism of the nonteleological doctrine of human behavior (we are moved exclusively by involuntary and self-centered drives) provoked a reaction in German Idealism. According to Immanuel Kant (1976, sec. 1), free will is restored, and with that a sense of virtue, as the resolve to do one's duty according to the moral law (the categorical imperative). Humanity can thus progress over time toward greater rationality; history becomes the domain of a new moral teleology in which individuals are respected as ends, not used as means. Nonhuman nature, however, remains "material to work on" (Locke, 1952, sec. 35).

Richard F. Hassing

See also Essentialism; Hobbes's Philosophical Method: Nature–Man–Society; Holism, in the Social Sciences; Institutions as Moral Persons; Kinds: Natural Kinds Versus Human Kinds; Laws of Nature; *Naturwissenschaften* Versus *Geisteswissenschaften;* Reduction and the Unity of Science; Vico's *Scienza Nuova*

Further Readings

Aquinas, T. (1948). *Summa Theologica* [Summary of theology] (Vols. 1 & 2, Question 92; Fathers of the English Dominican Province, Trans.). Allen, TX: Christian Classics.

Aquinas, T. (1986). *The division and methods of the sciences* (A. Maurer, Trans.). Toronto, Ontario, Canada: Pontifical Institute of Medieval Studies.

Aristotle. (1984). *The politics* (C. Lord, Trans.). Chicago, IL: University of Chicago Press.

Aristotle. (1995). *The physics* (J. Sachs, Trans.). New Brunswick, NJ: Rutgers University Press.

Aristotle. (1999). *Metaphysics* (J. Sachs, Trans.). Santa Fe, NM: Green Lion Press.

Aristotle. (2002). *Nicomachean ethics* (J. Sachs, Trans.). Newburyport, MA: Focus.

Bacon, F. (1960). *The new organon* (Fulton Anderson, Ed.). Indianapolis, IN: Bobbs-Merrill Library of Liberal Arts. (Original work published 1620 as *Novum Organum*)

Hassing, R. (1997a). Introduction. In R. Hassing (Ed.), *Final causality in nature and human affairs* (pp. 1–30). Washington, DC: Catholic University of America Press.

Hassing, R. (1997b). Modern natural science and the intelligibility of human experience. In R. Hassing (Ed.), *Final causality in nature and human affairs* (pp. 211–257). Washington, DC: Catholic University of America Press.

Hassing, R. (2008). Difficulties for natural law based on modern conceptions of nature. In A. M. Gonzalez (Ed.), *Contemporary perspectives on natural law* (pp. 229–240). Hampshire, England: Ashgate.

Kant, I. (1976). *Foundations of the metaphysics of morals* (L. W. Beck, Trans.). Indianapolis, IN: Bobbs-Merrill Library of Liberal Arts.

Kennington, R. (2004). *On modern origins: Essays in early modern philosophy* (P. Kraus & F. Hunt, Eds.). Lanham, MD: Lexington Books.

Locke, J. (1952). *The second treatise of government.* Indianapolis, IN: Bobbs-Merrill Library of Liberal Arts.
Machiavelli, N. (1985). *The prince* (H. C. Mansfield Jr., Trans.). Chicago, IL: University of Chicago Press.
Newton, I. (1969). *Pricipia: The mathematical principles of natural philosophy* (Preface and Book III, Proposition 7 and Corollary 2; A. Motte & F. Cajori, Trans.). New York, NY: Greenwood Press.
Spinoza, B. (1985). Ethics. In *The collected writings of Spinoza* (Vol. 1; E. Curley, Trans.). Princeton, NJ: Princeton University Press.
Strauss, L. (1968). An epilogue. In *Liberalism ancient and modern* (pp. 203–224). New York, NY: Basic Books. (Reprinted from *Essays on the scientific study of politics,* by H. Storing, Ed., 1962, New York, NY: Holt, Reinhart & Winston)

Legal Epistemology

Legal epistemology studies the relations between law and truth. To date, the bulk of work in legal epistemology has occurred in two areas. The first examines the ways in which Western criminal law systems, through their rules of evidence and procedure, standards and burdens of proof, and adversarial or inquisitorial forms of litigation, either facilitate or compromise truth finding in investigations of criminal wrongdoing. The second area covers debates over the truth status of propositions of law.

Law and Truth in Fact

All law addresses social situations that law purports to describe accurately as a first step toward regulating those social situations. For example, in deciding whether a plaintiff has had her interests damaged by a defendant in a way prohibited by law, courts must form a view regarding what is to count as the facts underlying the dispute. In deciding for either party, courts must rely on this picture of the facts when applying certain standards of proof to determine the likelihood or probability of a causal link between the defendant's action or inaction and the damage done to the plaintiff's person or property. In serving justice, courts must inevitably rely on a conception of truth, or at least a conception of what constitutes a reliable demonstration that a claim of fact ought to be treated as true.

Yet while all areas of law are committed in some fashion to truth, the most prominent work in legal epistemology is found in criminal law theory. Several features of criminal law attract analysis from legal epistemologists, precisely because they might either assist courts and juries in finding the truth about criminal wrongdoing or increase the chances of error. Exclusionary rules of evidence—for example, rules that prohibit the presentation of relevant and accurate evidence if the presentation of such evidence is likely to mislead juries, because juries are likely to overestimate its weight—are particularly interesting. Such rules might compromise truth finding by excluding relevant evidence or might, instead, indirectly better assist courts and juries in reaching true or warranted verdicts. Legal epistemologists also examine the level of a standard of proof for a finding of guilt and especially how such a standard of proof is explained in judicial instructions to juries. Still other aspects of criminal law of epistemological interest include the evaluation of expert testimony, the admissibility of confessions, the right of defendants and witnesses to remain silent, and the prevalence of adversarial forms of litigation. To settle many of the questions surrounding the success or failure of these features of criminal law, legal epistemologists are increasingly turning to complementary studies in the social sciences, especially those that investigate the social psychology and behavior of judges and juries.

Propositions of Law and Truth

Legal epistemology also refers to work on a central problem in the philosophy of law: What makes a proposition of law true? There are several, often competing answers that are typically associated with particular philosophical theories of knowledge.

According to legal *positivist* views, a proposition of law is true if it can be traced back to the foundational sources of law of a particular legal system. While positivists diverge in their characterizations of the foundational sources of law, nearly all agree that the foundational sources are rooted in observable social facts. For this reason, legal positivist views rely on empiricist types of epistemology, in which knowledge of law is grounded primarily in observation and experience. *Natural law* theories, by contrast, maintain that the truth of propositions of law is determined by objective moral standards. Such theories

therefore rely on teleological and objectivist moral epistemology, in which knowledge of law is acquired through moral reasoning about life's basic and ultimate goods. *Constructivist* theories occupy something of a middle ground between positivist theories and natural law theories. They posit an essential connection between the truth of propositions of law and moral principles but insist that the relevant moral principles are not to be found in some objective world of basic goods that exists outside and independently of the legal system. Instead, the relevant moral principles are to be found in the underlying purposes, aims, and rationales of existing judicial decisions, statutes, and constitutions. On constructivist views, a proposition of law is true if it forms part of the best holistic moral account of positive law.

Not all theories of law suppose, however, that the truth of propositions of law depends upon a correct view of rules, standards, or principles, whether legal or moral. Critical theories such as *legal realism* insist that the truth of propositions of law depends upon the dispositions, inclinations, and backgrounds of judges and courts. On this social-scientific view, law is inescapably shaped by historical, political, psychological, and economic factors, and so to understand law properly requires engaging in these and other social-scientific investigations. Later critical theories, such as feminist, critical race, and postmodern theories, have gone further and deny that propositions of law can be true at all; instead, propositions of law simply track the interests of the dominant class of society. Postmodern theories, for example, deny that there is truth in criminal judgments because these are merely expressions of state oppression.

Legal epistemology is by all accounts still in its infancy, and much work remains to be done in exploring how new work in epistemology might shed light on problems in legal theory beyond those surrounding criminal law and what makes a proposition of law true. These are exemplary areas of work but by no means exhaustive.

Michael Giudice

See also Epistemology; Law, Social Phenomenon of; Naturalized Epistemology; Social Epistemology; Truth, Philosophical Theories of

Further Readings

Burge-Hendrix, B. (2008). *Epistemic uncertainty and legal theory.* Farnham, England: Ashgate.

Giudice, M., Waluchow, W. J., & Del Mar, M. (Eds.). (2010). *The methodology of legal theory.* Farnham, England: Ashgate.

Haack, S. (2004). Epistemology legalized: Or, truth, justice, and the American way. *American Journal of Jurisprudence, 49,* 43–61.

Jackson, J., & Doran, S. (1996). Evidence. In D. Patterson (Ed.), *A companion to the philosophy of law and legal theory* (pp. 172–183). Oxford, England: Blackwell.

Laudan, L. (2006). *Truth, error, and criminal law: An essay in legal epistemology.* Cambridge, England: Cambridge University Press.

Patterson, D. (1996). *Law and truth.* Oxford, England: Oxford University Press.

Libertarianism, Metaphysical

Metaphysical libertarianism is a thesis about *free will.* It holds that free will is incompatible with determinism, and we have free will. This entry explains what is distinctive about this view of free will and describes several kinds of libertarian views.

Free will may be understood simply as what it takes to act freely—the requisite powers of recognizing reasons, deliberating, choosing, and executing one's intentions. When you act freely, you exercise free will; you perform a free action.

Philosophers employ different conceptions of free action. According to one of these, your action on some occasion was free, just in case you were able to do what you did and were able to do some other thing instead, and it was up to you which of these things you did. According to another conception, your action was free in case it can ground your moral responsibility for what you did on that occasion. While some philosophers think that these two conceptions of free action apply to exactly the same things, others think that an action can be free in the second respect without being free in the first respect.

Determinism is the thesis that the laws of nature are such that how the world is at any given point in time completely determines how it is at any subsequent time. If determinism is true, then a complete statement of the laws, together with a complete description of how the world is at any time, entails all the truths about how the world is at any later time. For example, if our world is deterministic, then given its laws, how the world was shortly after the big bang completely determines the fact that you

are reading this entry right now (see Pierre Simon Laplace's famous "demon" or, more accurately, "intelligence" so vast and powerful that by knowing the state of the whole universe at a certain moment can predict any of its future ones: "We may regard the present state of the universe as the effect of its past and the cause of its future").

To say that free will is incompatible with determinism is to say that if determinism is true, then free will cannot exist. In other words, if we have free will, it must be the case that determinism is false, that our world is indeterministic. The laws of nature must be such that there are some occurrences that are not completely determined by what happens prior to them. Libertarians typically think that free actions themselves—and particularly free choices—must be among these undetermined occurrences. Libertarians, then, deny *compatibilism,* which holds that we can have free will even if our world is deterministic (i.e., free will and causal determination are compatible).

What must a free choice be like according to libertarianism? For example, suppose that you deliberate about whether to tell the truth or to lie about some important matter, and you choose to tell the truth. What must your choosing be like for it to be the case that you freely choose to tell the truth?

Some libertarians hold that a free choice must be entirely uncaused, and it need not have any internal causal structure—it need not consist of one thing causing another. Two difficulties for this view concern how one can have any control at all over whether such an uncaused event occurs and how a choice can be made for reasons if one's deliberations do not cause that choice (these critical puzzles are even more evident when one tries to build social-scientific explanations of human actions in terms of a dichotomy between a world of freely acting human beings who control their social activities on the basis of deliberation and a natural world of causally determined physical events).

A second libertarian view holds that a free choice must be caused but not determined by prior events, including things such as the agent's appreciating various reasons for doing this or that. It is widely accepted that causal relations need not be deterministic. They can, for example, instantiate probabilistic laws. There could be a law that certain prior events will, with a probability of .6, cause an outcome of type E_1 or will, instead, with a probability of .4, cause an outcome of type E_2. When an agent's choice is caused but not determined, the laws of nature might allow that even with all prior events as they were the agent might have chosen otherwise. Still, a difficulty for this kind of view concerns whether it can really be up to the agent that she makes one choice rather than another.

A third kind of libertarian view invokes a special kind of causation—"agent causation"—that is said to be causation by the agent and not by any states of the agent (cognitive or motivational) or events involving the agent (her coming to have certain thoughts or motives). According to this kind of view, when an agent deliberates about what to do and freely chooses to do so, no states or events prior to the choice determine that choice, but the agent herself causes her coming to have some particular intention and thereby determines which choice she makes. Difficulties for this view include the task of explaining what this special kind of causation is supposed to be as well as providing reasons for thinking that it exists.

All these libertarian views face a challenge from science, for they all require that the laws of nature not be deterministic, that is, that our world be indeterministic. Some philosophers think that the well-confirmed status of quantum mechanics constitutes good evidence for indeterminism; but there are interpretations of quantum mechanics that are consistent with determinism, and the question remains open. Furthermore, libertarian views require that certain events, such as the making of choices, be undetermined by what precedes them. It is often thought that what is then required is that there be some undetermined neural events in our brains. Scientists have conducted some interesting experiments to find out whether this is so. As of yet, the verdict is uncertain.

Randolph Clarke

See also Action, Philosophical Theory of; Agency; Causation, Philosophical Views of; Causes Versus Reasons in Action Explanation; Determinism; Free Will, Philosophical Conceptions of

Further Readings

Clarke, R. (2003). *Libertarian theories of free will.* New York, NY: Oxford University Press.

Kane, R. (1996). *The significance of free will.* New York, NY: Oxford University Press.

Mele, A. R. (2006). *Free will and luck.* New York, NY: Oxford University Press.

O'Connor, T. (2000). *Persons and causes: The metaphysics of free will.* New York, NY: Oxford University Press.

Libertarianism, Political

Libertarianism is a theory of justice that holds that agents initially fully own themselves and have moral powers to acquire private property rights in external things. It judges nonconsensual force against a person to be just only when it is necessary to prevent that person from infringing someone's rights or to impose rectification for such infringement (e.g., compensation or punishment). These limits on the use of force radically limit the just powers of government.

Libertarianism is often thought of as "right-wing" doctrine. This, however, is mistaken for at least two reasons. First, on social—rather than economic—issues, libertarianism tends to be "left wing." It opposes laws that restrict consensual and private sexual relationships between adults, laws that restrict drug use, laws that impose religious views or practices on individuals, and compulsory military service. Second, in addition to the better-known version of libertarianism—*right libertarianism*—there is also a version known as *left libertarianism* (described below), which holds that unappropriated natural resources belong to everyone in some egalitarian manner.

Libertarianism can be understood as a basic principle or as a derivative one. Here, it will be understood as a basic moral/principle (i.e., based on natural rights). It is possible, however, to defend libertarianism as a derivative principle. Rule utilitarianism could lead to libertarian principles, as could rule contractarianism.

Libertarianism is normally advocated as a theory of justice in one of two senses. In one sense, justice is concerned with the *moral duties that we owe others*. It does not address *impersonal* duties (duties owed to no one) or *duties owed to self*. In a second sense, justice is concerned with the morally *enforceable* duties that we have. It does not address duties for which it is impermissible to use force to ensure compliance or to rectify noncompliance (e.g., a duty to see your mother on her birthday).

Libertarianism holds that agents are, at least initially, *full self-owners*. This means that they own themselves in just the same way that they can fully own inanimate objects. This full private ownership of a person or thing includes (a) *full control rights* over its *use,* (b) a *full power to transfer* these rights to others (by sale, rental, gift, or loan), (c) a *full right to compensation* if someone infringes these rights, (d) *full enforcement rights* to prevent infringement of these rights, and (e) *full immunity from the nonconsensual loss* of any of the rights of ownership, as long as one does not infringe the rights of others. The property rights in question are *moral* rights and may not be legally recognized.

Something like self-ownership is arguably needed to recognize the fact that there are some things (e.g., various forms of physical contact) that may not be done to a person without her consent but that may be done with that consent. One might, however, endorse full-*control* self-ownership without endorsing the other rights of full self-ownership.

Three main objections to full self-ownership are the following: (1) Because it holds that agents have not only the right to control the use of their person but also the right to *transfer* that right (e.g., by sale or gift) to others, voluntary enslavement is permitted; (2) full self-ownership entails that individuals have no duty of justice, except by voluntary agreement, to perform actions that help the needy; and (3) full self-ownership entails that forced service (e.g., draft into the military) is unjust, even when such service is needed to provide public goods. Libertarians typically defend voluntary slavery on the ground (roughly) that the right to *exercise* one's autonomy is more fundamental than the *protection or promotion* of one's autonomy. Libertarians typically defend the lack of a duty of justice to help the needy or to provide public goods on the ground that such duties involve a kind of partial involuntary slavery.

So far, we have considered agent self-ownership. There is an important distinction between *right libertarianism* and *left libertarianism,* depending on the stance taken on how natural resources can be owned.

Right libertarianism—the traditional form of libertarianism—holds that natural resources are initially nonowned and, typically, may be appropriated without the consent of or significant payment to others. It holds, for example, that whoever first discovers, first mixes her labor with, or first claims a natural resource owns that resource, provided, perhaps, that certain minimal conditions are satisfied. *Radical-right libertarians* hold that that there are no constraining conditions. Natural resources are

simply up for grabs. *Lockean right libertarians,* on the other hand, hold that appropriation is morally valid only if "enough and as good" is left for others (this is known as the Lockean proviso and is John Locke's position).

Left libertarianism, by contrast, holds that natural resources are owned by the members of society in some egalitarian sense, so that appropriation is legitimate only with their consent or with a significant payment to them. According to one version of left libertarianism, natural resources are *jointly owned* in the sense that authorization to use or to appropriate is given through some specified collective decision-making process (e.g., by majority or unanimous decision). The most well-developed and best-known form of left libertarianism is *Georgist left libertarianism* (as developed, e.g., by Henry George). It holds that agents may appropriate unappropriated natural resources as long as they pay for the competitive value (based on supply and demand) of the rights they claim. Equal-share versions divide up the rent pool equally among all, whereas "equal opportunity for well-being" versions divide up the rent pool unequally so as to equalize the opportunity for well-being.

Peter Vallentyne

See also Capitalism; Hayek and the "Use of Knowledge in Society"; Individualism, Methodological; Promises and Agreements; Public Goods; Self-Direction and Self-Ownership; Sen's Paretian Liberal; Spontaneous Order

Further Readings

George, H. (1966). *Progress and poverty* (5th ed.). Garden City, NY: Doubleday, Page. (Reprinted 1966, New York, NY: Robert Schalkenbach Foundation)

Locke, J. (1960). *Two treatises of government* (P. Laslett, Ed.). New York, NY: Cambridge University Press. (Original work published 1690)

Narveson, J. (1988). *The libertarian idea.* Philadelphia, PA: Temple University Press.

Nozick, R. (1974). *Anarchy, state, and utopia.* New York, NY: Basic Books.

Otsuka, M. (2003). *Libertarianism without inequality.* Oxford, England: Clarendon Press.

Rothbard, M. (1978). *For a new liberty, the libertarian manifesto* (Rev. ed.). New York, NY: Libertarian Review Foundation.

Life-World

The *life-world* is the world of immediate experience, existing prior to any organized or systematic form of objectification. When we think, feel, or perceive something or when we act, the life-world is presupposed as a horizon of certainties that are always already there, pregiven to us, and existing independently of active verification. Consisting of all the things we take for granted, it is the "ground" on which we stand as human beings. While structures of the life-world can be explored and revealed to us, it is impossible to grasp it as a totality or as some kind of complex object.

The concept of the life-world has played an important role both in philosophy and in the social sciences, as this entry will show.

The Concept of the Life-World in Philosophy

The concept of the life-world (*Lebenswelt*) has its origin in the works of Wilhelm Dilthey, who argued that knowledge is always situated in contexts of dynamic structures of meaning and perception. Prior to scientific activity, there is "life" (*Leben*)—hence the so-called Lebensphilosophie, which was highly influential in the latter half of the 19th century.

In the late works of Edmund Husserl, in particular *The Crisis of European Sciences and Transcendental Phenomenology,* the notion of the life-world attains crucial importance not only for understanding the formation of human knowledge but also for thinking about subjectivity, consciousness, experience, and meaning. Husserl makes various claims about the life-world. It is, first, the world as naively and immediately encountered, prior to any theorizing. In that sense, the life-world is the world of immediately encountered objects. However, it is also the world of such objects considered as use-objects or objects that may have some particular significance for us. Second, even when we do not think about them, the various features of the life-world are taken for granted. We unreflectively trust the life-world and do not expect any surprises. The life-world is essentially familiar to us. Third, although parts of it may be made explicit to us, we usually do not think much about the life-world. Fourth, Husserl compares the life-world to a horizon of certainties, suggesting that the horizon

must be there as a background for any individual objects, events, or tasks to emerge as meaningful. For example, the driving of a car makes sense to us because we presuppose all sorts of facts about the car and its physical environment, as well as about the traffic regulations and the possible actions of other drivers, which we rarely make explicit. Among other things, we presuppose that there is an engine in it, but we usually do not think about the engine unless it breaks down, needs service, and so on. Fifth, in Husserl's view, while the life-world exists independently of scientific idealization, science is grounded in the immediate experience of the life-world.

According to Husserl, the objectivating and idealizing physical sciences have tended to forget or refuse to acknowledge their dependence on the life-world. A crucial task of phenomenology is to recover these unacknowledged meaning-structures and make them relevant when thinking about subjectivity and man's place in the world.

Martin Heidegger, Maurice Merleau-Ponty, and other thinkers in the phenomenological tradition have used Husserl's account of the life-world as a foil for developing ontological theories of the human subject and its existence in the world. In part, through the body and its responses, the subject is viewed as profoundly situated in its environment—it has "being-in-the-world" as its most fundamental determination. The Cartesian notion that the subject and its mental states can be analyzed independently of the subject's dynamic relation to its surroundings is in this tradition rejected.

The Concept of the Life-World in the Social Sciences

The concept of the life-world has also been of great significance to the social sciences. In the work of Alfred Schütz and Thomas Luckmann, for example, the life-world is viewed as a repertoire of tacit knowledge not only of the physical surroundings but also of the social meanings available to the agent. Thus, an office space, a family home, or a factory is structured on the basis of numerous social codes and expectations the mastery of which is required for successful social coordination to be possible. Most of these codes are rarely made explicit to the relevant agents. They can be studied by the social scientist, yet ordinarily, they function as resources that the individual agent draws on in making sense of the situations at hand.

Jürgen Habermas has proposed that the concept of the life-world should be viewed as complementary to the concept of communicative action. According to Habermas, the life-world consists of a cultural, a social, and a personal component. Through communicative action—the making of speech acts—the life-world is both presupposed and reproduced. Habermas also claims that the life-world, especially in modernity, stands in danger of being "colonized" by processes originating in the economic and administrative systems. Thus, universities, for example, can be considered in terms of the notion of the life-world. Among other things, people in these institutions study, debate, and transmit cultural meaning. However, universities are also institutions that will have to respond to economic and administrative imperatives. In such cases, the life-world may be both supported and threatened by "the system."

According to some theorists, the life-world should not be theorized independently of social power and ideology. Thus, Pierre Bourdieu uses the term *habitus* to refer to the ways in which society at large, including its systems of power and ideology, is present in the individual and in the formation of individual identity.

The concept of the life-world has mainly been of importance to interpretive forms of social research. It tends to play a limited role in more formal approaches such as structuralism or system functionalism.

Espen Hammer

See also Agency; Being-in-the-World; Existential Phenomenology and the Social Sciences; Habitus; Hermeneutics, Phenomenology, and Meaning; Knowing-How Versus Knowing-That; *Naturwissenschaften* Versus *Geisteswissenschaften;* Phenomenological Schools of Psychology; Social Construction of Reality; Tacit Knowledge

Further Readings

Bourdieu, P. (1992). *The logic of practice.* Stanford, CA: Stanford University Press.

Dilthey, W. (1989). *Introduction to the human sciences: An attempt to lay a foundation for the study of society and history.* Detroit, MI: Wayne State University Press.

Habermas, J. (1985). *The theory of communicative action: Vol. 2. Lifeworld and system: A critique of functionalist reason.* Boston, MA: Beacon Press.
Heidegger, M. (2000). *Being and time.* London, England: Blackwell.
Husserl, M. (1970). *The crisis of European sciences and transcendental phenomenology.* Evanston, IL: Northwestern University Press.
Merleau-Ponty, M. (2002). *Phenomenology of perception.* London, England: Routledge.
Schütz, A., & Luckmann, T. (1973). *The structures of the life-world* (2 vols.). Evanston, IL: Northwestern University Press.

Logical Positivism/Logical Empiricism

This entry gives an overview of the once influential philosophical movement of the Logical Positivists or Logical Empiricists, among the main contributors who shaped 20th-century analytic philosophy and philosophy of science in particular. The history of the philosophy of the social sciences has been particularly entangled with the fortunes of the logical-positivist views of science. The entry situates the movement within classical positivism, presents its main doctrines, and recounts its rise and fall as well as its current echoes, among which one may still discern a philosophical predilection in favor of science.

Background

The terms *logical positivism* and *logical empiricism* refer to the movement of scientific philosophy that originated in the late 1920s with the Vienna Circle and the Berlin Circle, had a major impact on philosophy in America and elsewhere, and was widely abandoned in the 1960s. The two labels are often used synonymously (as in this entry). Some authors employ the former to refer to the Vienna Circle and the latter to the Berlin Circle, while others apply them, respectively, to the early phase and the later phase of the movement, after its migration, mainly to the United States, during the Nazi era.

Following the assassination of the philosopher Moritz Schlick in 1936, the mathematical logician Rudolf Carnap and the social scientist Otto Neurath became the leaders of the Vienna Circle, while the physicist-philosopher Hans Reichenbach headed the Berlin group. Recent scholarship has disclosed the logical empiricist movement to be far more diverse and interesting than previously portrayed. The founders had advanced scientific training, but their interests differed. A full account of the various streams of the movement is beyond the scope of a brief encyclopedia entry. The movement's overall aim, which invited accusations of scientism, was to continue the project of the scientific Enlightenment by bringing all aspects of culture—from politics to art and architecture—in line with modern science as the most reliable and progressive human endeavor.

Logical Empiricism in the Positivist Tradition

Positivism is a family of strongly empiricist philosophies that anchor inquiry in facts established by observation as the basic units of scientific achievement. Theoretical claims are suspect insofar as they cannot be appropriately anchored in the facts. Positivists have always regarded human culture, not only including political and religious debate but also science, as being needlessly burdened by turgid metaphysical language that obfuscates and hence promotes ignorance. Accordingly, the most general aims of positivist movements have been to purify culture, to make human communication and decision making transparent across all social levels and thereby more rational, and to satisfy the human need to know by replacing outdated religious and other metaphysical ideas by modern science as providing our best account of our place in the universe.

The logical positivists/logical empiricists formulated a positivism that combines an empiricist epistemology with the resources of the new symbolic logic of Bertrand Russell and Alfred North Whitehead, given the latter's remarkably increased richness of precise symbolic expression and deductive power. For them, science is best characterized by its method, and in Carnap's influential view, method is to be understood as the logic of science. Nineteenth-century methodologists had used similar language, but now "logic" meant the new formal logic, not the sometimes woolly, idealist disquisitions from the past. Meanwhile, Neurath dreamed of a new *Encyclopedia of Unified Science* as the primary

cultural expression of the movement, just as the 18th-century French *Encyclopédie* had epitomized the original scientific Enlightenment. Relatively few installments were published, the last of which, Thomas Kuhn's *Structure of Scientific Revolutions,* in 1962, was widely perceived as marking the end of the movement. Later scholarship, however, has revealed a surprising degree of continuity, especially in Kuhn's treatment of scientific language and in his quasi-Kantian, two-tier conception of science, in which a stable layer or framework of constitutive rules or understandings makes possible the normal, esoteric work of everyday research. Reichenbach and Carnap held somewhat similar views prior to Kuhn, although Kuhn emphasized that even mature sciences will, of necessity, undergo revolutionary transformations from time to time. This residue of Kantianism distinguishes the Germanic empiricism of leading logical empiricists from that of the English empiricist tradition from John Locke to Bertrand Russell and A. J. Ayer.

Characteristic Doctrines

Attributing a collection of theses to the logical empiricists runs the risk of underplaying the internal diversity of the movement, but it is safe to say that the following themes were central to their scientific world picture:

a. *The unity of science, with physics as the model science:* All sciences can be reconstructed as employing the method of physics and as based on the same pure observation language, a "physical thing language" based on easily detectable properties and relations of physical objects. Thus, according to Carl Hempel and contrary to the *Verstehen* or historical understanding tradition, historical explanations are good only insofar as they can be regarded as sketches of deductions from law statements, like those found in physics.

b. *Physicalism, reducibility to physics:* All scientific claims can be given a physical meaning, and at least some theories or sciences are reducible to physics as the basic science—perhaps the entire hierarchy, from the social sciences through psychology, biology, and chemistry.

c. *An empiricist theory of meaning:* Theoretical statements are admissible if and only if they are empirically verifiable or at least rigorously testable against experience, since they derive their meaning from connections to the observation language.

d. *Rejection of metaphysics:* Any claim that pretends to assert something about the world but fails the above criterion is meaningless nonsense.

e. *A sharp fact/value distinction:* Value claims such as those of ethics and aesthetics are emotional expressions, not factual claims.

f. *Descriptionism or instrumentalism versus theoretical realism:* Since theoretical language is inherently problematic, science should bridge the traditional gap between appearance and reality by regarding theories as descriptive reorganizations of bodies of experimental facts rather than as representing a reality behind the appearances; in other words, theories should be seen merely as instruments for calculating predictions.

g. *Reconstructive purity:* Whatever is worth saving in theoretical science can be reconstructed transparently in terms of formal logical structures satisfying the above empiricist criterion of meaning, hence the goal of providing a general, structural account of theories, confirmation, probabilistic inference, prediction, explanation, theory reduction, and so on.

h. *The distinction between context of discovery and context of justification:* There exists a logic of confirmation but no logic of discovery. Thus, the process of research, and the history of science more generally, can be ignored. Only the logical structure of the finished product is relevant to epistemology.

i. *The normative/descriptive distinction:* Philosophy (including philosophy of science) is a normative discipline, not merely descriptive.

j. *The context freedom and hence objectivity of scientific claims:* The logic of science and the empirical facts together provide a strong directive to research, one that is immune to the disputes over interpretation and the biases and

contextual relativity that characterize the humanities and political and religious discourse.

k. *Value freedom:* Science itself is value-free. While crucially important to an enlightened society, it provides no particular political agenda.

l. *Science and scientific philosophy are construed as cooperative, incremental, cumulative, and progressive enterprises:* The linguistic and methodological unity of science enables the sciences, unlike other human endeavors, to make steady progress, adding to the storehouse of scientific knowledge, always within the same, gradually expanding conceptual framework. Technical philosophy can emulate this model.

The sometimes crazy politics and social movements of the Weimar period motivated the logical empiricists' language reform program. A related, democratic motivation was to make scientific language so transparent as to be intelligible even to the working class. They struggled to solve the problem that logic, mathematics, and hence philosophy were not, in their view, empirically testable yet were indispensable. Clues from Russell and Ludwig Wittgenstein inspired their solution to the problem: These statements do not need to satisfy the criterion because they make no claims about the world. They are empty of empirical content but in a good way. Their function is to provide inferential structure only. Writ large, this implied that philosophy, as the logical analysis of the language of science, can be scientific without itself being an empirical science.

When logical empiricism came to America, it fairly quickly achieved a degree of hegemony, displacing the then reigning pragmatism of William James and, especially, John Dewey. Pragmatism was fairly congenial to positivism but seemed imprecise and out of date by comparison, and the central logical empiricist theses did hang together well. Of course, highly integrated approaches are robust in some respects but vulnerable in others in that serious trouble anywhere can propagate quickly through the entire system. That is ultimately what happened to logical empiricism.

The Fall of Logical Empiricism

The logical empiricists were their own worst critics. Already by 1950, Hempel could argue that attempts to formulate a logical criterion of cognitive significance were doomed. He added that concept formation cannot be done in isolation as the operationists insisted, that it is intimately coupled with active theorizing. About the same time, the American logician and philosopher Willard Van Orman Quine parted intellectually from Carnap by rejecting the so-called analytic/synthetic distinction, thereby denying that logic and mathematics have a special, nonempirical status, forever immune to revision. Soon, others began to challenge the observational/theoretical distinction and the idea that human learning starts from bare experience and proceeds only gradually up the empiricist ladder of increasing grades of theoreticity. Moreover, it was becoming apparent that theory and experimental observation are often tightly coupled and that many observations are scientifically significant only in light of some theory. The discovery/justification distinction also came under attack; for the new history of science suggested that the research process (i.e., discovery) could be studied in epistemologically interesting ways and that focusing solely on the logical structure of finished products (i.e., on justification of existing theories), as orthodox logical positivism demanded, produced only formalizations of textbook science and failed to capture real science as practiced. To analysts such as Stephen Toulmin, Kuhn, Paul Feyerabend, Imre Lakatos, and Larry Laudan, science does not escape from history but is itself subject to dynamic changes that cannot be captured by static, "time-slice" logical structures. The disparity between the positivist model of science (including the somewhat similar views of Karl Popper) and what historians were disclosing was so great that it was not enough for logical empiricists to fall back on the defense that philosophy of science is normative rather than descriptive (i.e., it dictates what correct scientific reasoning and justification should be rather than describes actual scientific practice). Kuhn and Feyerabend especially attacked the view that science is cumulative over historical time, always within the same, albeit expanding, conceptual framework. On the contrary, Kuhn said, even the most mature sciences will occasionally and unpredictably undergo transformative revolutions that change the rules of the game. Logic and facts do not suffice to determine major scientific decisions. Alternative interpretations and rhetorical persuasion are possible at every stage but are normally suppressed. Revolutions do

not result from large inputs of new facts but from wholesale conceptual reinterpretations of established results already in hand. As the 1960s and 1970s wore on, it became clear to most parties that logical empiricism lacked the resources to respond adequately to these criticisms. That the movement was judged to be wrong about science itself did not bode well for its original program to reshape modern culture, which, anyway, the logical empiricists abandoned in the conservative American political context.

Nonetheless, the logical empiricists left a legacy of high standards of clarity and rigor that have been fundamental to analytic philosophy. They also shaped the discipline of philosophy of science in a manner recognizable today. Although most philosophers now reject the logical empiricist positions, they often engage the same problematic. Thus, the unity of science remains a central issue, although many writers now reject unity in favor of diversity. Scientific realism remains a central topic, and some philosophers persist in the attempt to develop general theories of confirmation, explanation, and so on. Philosophers of science still tend to have a pro-attitude toward the sciences and to be more normative than other science studies experts, who strive to achieve a stance of descriptive neutrality.

Thomas Nickles

See also Analytic/Synthetic Distinction; Context of Discovery Versus Context of Justification; Empiricism; Instrumentalism of Scientific Theories and Constructive Empiricism; Kuhn on Scientific Revolutions and Incommensurability; *Naturwissenschaften* Versus *Geisteswissenschaften;* Neurath's Unity of Science and the Encyclopedia Project; Positivism, History of

Further Readings

Cartwright, N., Cat, J., Fleck, L., & Übel, T. (Eds.). (1996). *Otto Neurath: Philosophy between science and politics.* Cambridge, England: Cambridge University Press.

Friedman, M. (1999). *Reconsidering logical positivism.* Cambridge, England: Cambridge University Press.

Giere, R., & Richardson, A. (Eds.). (1996). *Origins of logical empiricism.* Minneapolis: University of Minnesota Press.

Hardcastle, G., & Richardson, A. (Eds.). (2003). *Logical empiricism in North America.* Minneapolis: University of Minnesota Press.

Richardson, A., & Übel, T. (Eds.). (2007). *The Cambridge companion to logical empiricism.* New York, NY: Cambridge University Press.

Love, in Social Theory

This entry presents the changing conceptual and social frameworks within which relations of love must now be conceptualized in conditions of cosmopolitanism and globalization. It also explains the emergence of the novel social phenomena of "long-distance love" and "world family" and sets the tasks of social sciences vis-à-vis these novel social relations of intimacy.

Introduction

We are living in an era of migration and globalization. Yet until now, research on love and the family has paid scant attention to this. Instead, it continues to focus mostly on the forms of personal life and the relationships characteristic of the majority society. Thus, it remains prisoner to the unholy trinity of territory, state, and nation. In other words, it remains trapped in "methodological nationalism"—that is, in a frame of reference that equates society as such with nationally organized societies. Yet today, this frame of reference is rapidly becoming anachronistic. It cannot deal with the emergence of ever more forms of personal life and relationships that extend across (national) borders.

Changing Frameworks

The longer family research clings to the "container model" of the nation-state, the less it will be able to understand and explain the situation of long-distance love and families in today's cosmopolitanized world. What is needed is a *cosmopolitan turn* in research on love and families. But what does cosmopolitanism mean?

Cosmopolitanism is a very old and highly contested set of ethical and political ideas, dating back to Hellenistic (Stoic) and Roman antiquity, combining—as the word *cosmo-polis* says—two worlds: Every human being is, first, a member of the *cosmos* (the unity of nature and humanity) and, second, at the same time, a member of the *polis*—that is, different states, ethnicities, and religions.

This idea was revived by the philosophies of the European enlightenment, most notably by Immanuel Kant, who foretold a future era of polite civilization, commerce, and global peace. By the 20th century, however, cosmopolitanism had virtually disappeared as a serious intellectual, let alone political, position.

"Great powers"–based realism instead dominated the mood of politics and the mode of analysis in an era defined by raging nationalisms, the trauma of world wars, and a Cold War in which competing superpower "internationalisms" vied for global dominance.

With the emergence of *globalization* as a master concept in the social sciences, cosmopolitanism regained currency within the academy. But in the process, cosmopolitanism also refashioned itself, moving beyond political theory, its conventional home, and ranging widely across anthropology, cultural studies, literary criticism, legal studies, and social history. New, more or less reflexive cosmopolitanisms have since proliferated, preoccupied, first, with squaring the circle of abstract universalism by emphasizing respect for the particularity of human diversity and, second, with expanding the boundaries of the circle to include (if not to favor) those for whom cosmopolitanism is not a lifestyle choice but the tragic involuntary condition of the refugee or the otherwise dispossessed.

Changes in Conceptualizing "Love"

What changes are love and family undergoing in the "global age"? The belief in the either/or, which was once taken to be a "natural" self-evident fact—either we or they, either here or there—seems to be on the wane or has actually disappeared from the horizon of love. Nothing now seems to separate human beings any longer in any absolute way, not skin color, not national hostility, not religious differences, not the distance between continents, and so forth. On the contrary, people are susceptible to the attractions, even the lure, of the unlimited possibilities represented by the global other—by those who are far away: *long-distance love*. This marks the disappearance of unbridgeable chasms; for since they now appear bridgeable, they are already in the process of being bridged.

To begin with, we have to distinguish between the social model of *national* families and love and the social model of *world* families and *long-distance* love. The vantage point that enables us to gauge the dimensions of the new landscapes opening up for love, family, and the household in the global age is as follows: The national or territorial face-to-face model of the "family"—defined by the unholy trinity of territory, passport, and skin color—is now breaking down.

The "world families" and "long-distance love" model, by contrast, includes a variety of forms in which the members/partners coexist across national, geographical, and religious boundaries—at the one extreme, different native cultures coexist in a single place; at the other extreme, a single native culture is scattered across the world (and between these extremes there are many transitional and hybrid forms). Including these different forms of the global other in one's own life calls for active trust from all sides across institutionalized boundaries. Only thus can the world's antagonisms be contained under the roof of the family or be rendered useful for familial purposes. The territorial or national family and love model is founded on the fact that the modern dualisms—national and international, global and local, we and the others—are accepted as "natural"; by contrast, the world families and the long-distance love model rests on exactly the opposite fact, namely, that these dualisms are placed in question in the practice of coexistence beyond boundaries.

Long-distance love and world families, including binational couples, marriage migrants, fertility tourists, foreign domestic workers and transnational households, "global care chains," and so on, are no longer marginal phenomena; they have long since taken root at the heart of the "majority society." "The global 'other' lives in our midst" acquires here a literal, intimate, and familial connotation. One's brother-in-law now has a wife from Thailand. A woman from Poland has been hired to look after grandpa. One's godchild has recently started living with a theologian from Togo. Where is Togo, actually? How come he is here? Is he here for the sake of a residence permit or out of "true" love?

World families integrate a "global other"—who as a result is no longer a global other. To bridge the institutionalized national and religious stereotypes and enemy images of the globe in everyday family life requires *active trust* from both sides in order to open up a passage for hope, by acknowledging being equal and different at the same time.

This vision might then become a source of disappointment and failure. But it is necessary because world families embody the contradictions of the world, and these contradictions are worked out in them. Not all families embody all contradictions, but some families embody some of them. For example, there are marriages, parents, and couples with dual nationality, and they may embody the tensions between two countries or between the majority and

minority communities within those countries, while immigrant families may incorporate the tensions between the center and the periphery or between legal and illegal members.

World families and long-distance relationships mirror a state of ignorance that has been nationally programmed and embodied in law. It follows that love and the family become the setting in which the "cultural wounds"—the rage and the anger that global inequalities and their imperial history continue to inspire in the souls of the living to this day—are endured and fought out.

World families, therefore, are not families with global *power* or families with global *horizons;* nor are they *one*-world families or families of world *citizens*. We might instead call them families embodying *world conflicts,* families ripe for *world adventures,* or even families *seeking their fortunes* in the world—families attempting to turn poverty and conflict into "gold." But what they demonstrate is that the universal image of the "good family," which has always been taken for granted, is now fundamentally in flux—a development that bears some responsibility for fundamentalist reactions and countermovements.

Social Science's Task

In order to study global intimacy, long distance love, and world families—that is, to bring the social sciences in—there is an urgent need to clearly distinguish *cosmopolitanism* in the normative sense of philosophy and political theory from *cosmopolitization* as a social-scientific concept and a program for empirical research. Cosmopolitanism is about *norms,* cosmopolitization is about *facts*. The argument is that the interrelatedness of people and of populations around the globe can only be understood from a cosmopolitan standpoint. Building this cosmopolitan approach implies the erosion of distinct boundaries dividing markets, states, civilizations, cultures, and, not the least, all life-worlds of different peoples. This is true not only for the study of love and family relations but also for the study of work conditions, religions, classes, nation-states, climate change, and so on.

Cosmopolitization does *not* mean that individuals living in "world families" are becoming cosmopolitan. We have to make a clear distinction between the perspective of the actor and the perspective of the social-scientific observer. The word *cosmopolitan* becomes indispensable for describing a situation in which—not only in world families but also in relation to world religions or global risks—*humanity* and *world* are not merely thinkable but unavoidable categories for describing the realities we live in—a "cosmopolitan situation," resulting in a growing *consciousness* of both the world as a single place and humanity as a unity of (often contradictory) differences. Culturally speaking, then, cosmopolitization has gained significance in a set of collective existential dilemmas at the start of the third millennium. Thus, with the empirical phenomena of "long-distance love" and "world families," we are witnessing the opening of a new chapter full of strife and controversy over the very meaning of "romantic love," "good family," or, more generally, the "good world"—and, significantly, who gets to define it.

Ulrich Beck and Elisabeth Beck-Gernsheim

See also Collective Emotions; Emotions; Love, Philosophy of; Personal Identity and Trauma; Postmodernism; Sexuality

Further Readings

Albrow, M. (1996). *The global age: State and society beyond modernity.* Cambridge, England: Polity Press and Blackwell.

Beck, U. (2006). *The cosmopolitan vision.* Cambridge, England: Polity Press.

Beck, U., & Beck-Gernsheim, E. (2011). *Fernliebe: Lebensformen im globalen Zeitalter* [Long-distance love: Forms of life in the global era]. Berlin, Germany: Suhrkamp.

Beck, U., & Grande, E. (2010). Varieties of second modernity: The cosmopolitan turn in social and political theory and research. *British Journal of Sociology, 61*(3), 409–443.

Beck, U., & Sznaider, N. (2006). Unpacking cosmopolitanism for the social sciences: A research agenda. *British Journal of Sociology, 57*(1), 1–23.

Hochschild, A. R. (2000). Global care chains and emotional surplus value. In W. Hutton & A. Giddens (Eds.), *On the edge: Living with global capitalism* (pp. 130–146). London, England: Jonathan Cape.

Wimmer, A., & Schiller, N. G. (2002). Methodological nationalism and beyond: Nation-state building, migration and the social sciences. *Global Networks, 2*(4), 301–334.

Yates, J. J. (2009). Mapping the good world: The new cosmopolitans and our changing world picture. *The Hedgehog Review, 11*(3), 7–27.

Love, Philosophy of

Love has attracted philosophical interest from antiquity to the present. Analytically speaking, love has been divided into many types: romantic love, parental love, and sexual love; love for friends, religion, country, food; and so on. This entry focuses on *romantic love,* which appears to be one of the most complex of emotions. It gives an overview of various philosophical analyses of its features and presents points of dispute about its nature.

The experience of romantic love involves two basic evaluative patterns, referring to (1) attractiveness, an attraction to external appearance, and (2) praiseworthiness, positively appraising personal characteristics. Falling and staying in love requires the presence of both patterns. These patterns are not independent: A positive appraisal of your partner's characteristics is greatly influenced by his or her attractiveness. A common phenomenon in romantic relationships is the "attractiveness halo," in which attractiveness positively influences ratings of intelligence, sociality, and morality.

In contrast to romantic love, where both evaluative patterns are essential, in sexual desire attraction is far more dominant. Sexual desire is a simpler attitude, based largely on spontaneous evaluations, whereas romantic love often requires both spontaneous and deliberative evaluations. In friendship, which is a kind of companionate love, praiseworthiness is by far more dominant than attractiveness.

Change is of greater weight in generating sexual desire; conversely, friendship depends more on familiarity and shared history. Since romantic love consists of both patterns, it is often difficult to maintain romantic relationships over the course of many years. In some cases, romantic love can remain intense for many years; in others, it becomes companionate love, in which the passionate aspect is reduced; and in yet other cases, the relationship does not survive in the long term. This raises doubts concerning whether, as in the popular song of the mid 20th century, "love and marriage go together like a horse and carriage."

A few classical disputes concerning the nature of romantic love are on (a) love at first sight, (b) whether love is blind, (c) the importance of similarities and differences in generating love, and (d) the exclusiveness of love.

a. Despite its glamorous place in our culture, love at first sight is not easy to explain. How can we explain that after one quick glance we can fall profoundly in love? Although people often confuse love at first sight with sexual desire at first sight, there are nevertheless genuine cases of love at first sight. The fundamental mistake in denying the existence of such love is the assumption that we cannot attribute to a person characteristics that are not present at the moment. Praiseworthiness can be evaluated spontaneously, much like attractiveness. This is done in many types of stereotypic evaluations, and it is also present in the aforementioned "attractiveness halo," in which what is beautiful is also evaluated as good. Accordingly, attractive people are more likely to be the object of love at first sight.

   There is, however, no need to adopt Shakespeare's assumption that "whoever loves, loves at first sight." Love can develop over time. We may discover characteristics that we adore but that were not evident at first sight, and we may evaluate differently those characteristics that at the beginning were evident or were not positive enough to make us fall in love. Hence, we can fall in love with people once we get to know them better. Sometimes, "to love him is to know him."

b. Lovers in general, but particularly lovers at first sight, are often blind to the beloved's negative traits. We frequently love the idealized person rather than the real one. Idealization of the beloved may also be considered as a kind of a defense mechanism, enabling us to justify our relatively arbitrary choice. The lover's blindness is not necessarily due to misperception of the beloved's traits; it may also be a matter of emphasis (or overemphasis), namely, focusing upon positive qualities only.

   To persist, an idealized loving relationship requires a moderate but constantly distorted perception of reality; its moderate nature enables lovers to believe in it despite its known inaccuracies. Stable and satisfying relationships

reflect the intimate partners' ability to see their imperfect partner in an idealized light. Love is thus not blind, but its sight is often blurred.

c. Concerning the classical dispute of whether "birds of a feather flock together" or "opposites attract," there is consistent evidence for similarity. Romantic partners show (i) strong similarity in age and in political and religious attitudes; (ii) moderate similarity in education, general intelligence, and values; and (iii) little or no similarity in personality characteristics. We tend to fall in love and stay with people who share with us profound similarities; the presence of such essential similarities need not preclude—and may even encourage—differences related to surface manifestations within the basic similarity. Differences attract, but only within a shared general framework that leaves ample space for complementary differences.

d. Romantic love requires components of both exclusiveness and uniqueness. Exclusiveness is characterized in negative terms, such as "not permitting" and "restricting," that establish rigid boundaries, like not being intimate with other people. Uniqueness is characterized in positive terms that celebrate an ideal and establish distinctiveness: "being made for each other." While romantic love involves both components, uniqueness is of greater significance for a profound, long-term romantic relationship.

Romantic love is frequently characterized in ideal terms, such as being comprehensive (there are no boundaries to this love), uncompromising (nothing can dilute or impede such a love), and unconditional (reality is almost irrelevant to love and has scant impact on it). Ideal love is perceived as an overwhelming force that can cope with all obstacles, not merely in the sense that it can solve all difficulties but also in the more profound sense that such difficulties, even if they continue to exist, become less important. The assumption that love can conquer all gives no weight to changing personal and contextual features and can be a source of disappointment and distress, as people might blame themselves or their beloved for having a relationship that is short of perfect.

Like other types of relationships, romantic love involves various compromises, such as (a) compromising on the overall value of the person, (b) compromising on the value of the person as a partner, and (c) compromising on the nonromantic activities within the given framework of their togetherness. In cases of profound, long-term romantic love, the feeling of compromising scarcely exists.

Our current era can be considered as the best and worst of times for lovers. This is indeed both a happy and a difficult time for relationships—happy in that available and willing potential lovers are all around but difficult in that maintaining a loving and committed relationship is harder than before since alternative romantic options have become easier to explore and to realize. The need to make romantic compromises is therefore greater, but making them is more difficult.

Aaron Ben-Ze'ev

See also Emotions; Love, in Social Theory; Sexuality

Further Readings

Ben-Ze'ev, A. (2004). *Love online: Emotions on the Internet.* Cambridge, England: Cambridge University Press.

Ben-Ze'ev, A., & Goussinsky, R. (2008). *In the name of love: Romantic ideology and its victims.* Oxford, England: Oxford University Press.

Fisher, H. (2004). *Why we love?* New York, NY: Holt.

Kipnis, L. (2003). *Against love: A polemic.* New York, NY: Pantheon.

Schnarch, D. (1997). *Passionate marriage: Love, sex, and intimacy in emotionally committed relationships.* New York, NY: W. W. Norton.

Singer, I. (1984a). *The nature of love I: Plato to Luther.* Chicago, IL: University of Chicago Press.

Singer, I. (1984b). *The nature of love II: Courtly and romantic.* Chicago, IL: University of Chicago Press.

Singer, I. (1989). *The nature of love III: The modern world.* Chicago, IL: University of Chicago Press.

Soble, A. (1999). *Eros, agape, and philia.* St. Paul, MN: Paragon House.

Soble, A. (2008). *The philosophy of sex and love: An introduction* (2nd ed.). St. Paul, MN: Paragon House.

Luhmann's Social Theory

The work of Niklas Luhmann occupied a central, albeit paradoxical, position in debates in the later 20th century regarding the position of philosophy vis-à-vis the social sciences. This entry presents an overview of Luhmann's social theory, his views on the relation between philosophy and social science,

and the importance of his theory for social-scientific epistemological issues.

In many respects, Luhmann's theory was conceived as a frontal assault on philosophy. In his inaugural lecture of 1967, titled "Sociological Enlightenment," he dismissed the residues of post-Enlightenment subjectivism and philosophical ontology in the social sciences. Indeed, he declared a programmatic intention to develop his sociology—or his *theory of society*—as a line of inquiry defined by a rejection of philosophical epistemologies and committed to strategies of positivistic/functionalist social analysis without reliance on implicitly metaphysical concepts of rationality, reflexive norm production, or cognitive monadicism.

In this respect, Luhmann positioned his thought around a sharp dichotomy between sociology and philosophy. He defined philosophy as the study of the human mind, and he construed sociology, strictly, as the study of society. In making this distinction, Luhmann proposed an *apersonal and anti-anthropological* conception of society, which defined society as a densely contingent repository of *social communications*. The communications forming society, for Luhmann, are not in any determinate way *human* communications. They are in fact *functional* or *systemic* communications: That is, they are communications within the various function systems that society comprises. The primary *function systems* in society are economy, art, science, law, politics, medicine, media, religion, and education. For Luhmann, each system possesses a distinct functional rationality, and through this rationality, it constructs a realm of differentiated functional meaning, which is in itself entirely self-referential, self-reproductive, or *autopoietic*. For example, those exchanges in society that are communicated as law (prosecutions, hearings, jury appointments, draft statutes, white papers, etc.) are communicated in categories that are exclusively internal to the functions of the legal system. Such exchanges are articulated for society around the code lawful/nonlawful. Similarly, those exchanges in society that are communicated as economics (e.g., transactions, loans, investments, warranties, taxation, etc.) are communicated in categories that have meaning solely for the functions of the economic system. Such exchanges are expressed for society around the code payment/nonpayment. As social meaning is communicated in differentiated systems, there is no overarching human rationality that encompasses all communications in society or reaches across systemically differentiated communications. There are in fact *multiple rationalities* in society, each inhering in a given social system, and each of these rationalities constructs inner-systemic meanings in indifference to any originally human rational substrate.

This systemic-communicative account of society underpinned Luhmann's attitude to philosophy. Like Émile Durkheim, he concluded that to approach societal objects from the presumption that they are interpretable by a unifying human rationality is to ignore the radical fact of functional and communicative differentiation, which is foundational for modern society. On this basis, furthermore, he concluded that the methodological preconditions of philosophy are inherently false and that in positing generalized principles of rationality, philosophy obstructs understanding of phenomena as specifically societal constructs. It is specific to sociology, however, that, *contra* philosophy, it can accept high levels of differentiated contingency, multirationality, communicative self-reference, and systemic internality as preconditions for the emergence of societal phenomena. Sociology, in other words, has the methodological distinction of being capable of interpreting objects *postontologically*—as phenomena arising as articulations of highly differentiated, and fundamentally abstracted, communicative contingency.

As an extension of this radical-functionalist approach, Luhmann concluded that the most basic assumption of philosophy—namely, that there exists a perennially constructed human intelligence capable of expressing itself as *theory*—is misguided. Against the classical idea of *theory,* he asserted a concept of *semantics*. The concept of semantics implies that each social system has an internal theoretical structure, and it consolidates its internal rationality by generating distinct theoretical concepts (semantics) to secure and temporally to stabilize the meanings specific to it. In consequence, the theoretical models emerging in different realms of social practice need to be examined, not—philosophically—as autonomous expressions of intelligence but rather—sociologically—as objectivized *self-descriptions* of particular social systems. Against the entire tradition of post-Platonist philosophical reflection, therefore, Luhmann concluded that the ideal structure of reason and the material structure of society need to be viewed as in essence *communicatively identical.*

Luhmann's ultracontingent conceptions of theory and rationality brought him into repeated conflict with intellectuals in the philosophical mainstream. For example, these constructions formed the basis of high-profile controversies between Luhmann and Jürgen Habermas, which punctuated West German intellectual life in the 1970s and 1980s. In these exchanges, Luhmann argued for a concept of theoretical communication as absolutely contingent meaning. Habermas rebutted this from a remotely post-Kantian perspective, defining theoretical communication as a bearer of anthropologically constitutive rational norms. Central to this dispute was the fact that Luhmann dismissed Habermas's normative sociology as a belated example of archaic metaphysical-ontological thinking, which was incapable of true sociological comprehension.

Although he positioned his theory of society as a denunciation of the ontological residues of philosophy, Luhmann's work was formed through interaction, albeit dialectical, with the philosophical canon. His account of sociology as variable interpretation has a certain proximity to the historicist-hermeneutical tradition. Most strikingly, however, Luhmann's view of society reposes on a philosophical construction of *difference,* in contrast to identity, as the basic substructure for social meaning. That is, Luhmann argued that meaning is generated in (or *as*) society through the endless differentiation of one social system from its environment (other systems) and that each act of systemic self-differentiation constitutes meaning as a contingent negation of other meanings (meaning constructed through the rationality of other systems). In this regard, Luhmann's thought discloses a distant affinity to Hegelian phenomenology.

More palpably, however, his theory of society was organized as a sociological reconstruction of Martin Heidegger's critique of post-Cartesian ontology, and it transposed the *Sein/Dasein-Differenz* in Heidegger's work into a sociological theory of systemic differentiation and contingent social meaning. For Luhmann, as for Heidegger, intelligible meaning is the result of societal communications defining themselves, contingently, in incessant and *ex-static* inner-societal difference from all other meanings and from all perennial or universally founded intelligence. Like Heidegger, Luhmann saw his own work as containing a method for capturing such meanings. To this extent, Luhmann appears to have viewed his theory of society as the culminating position in the radically temporalizing and radically pluralistic critique of formal ontology first promoted by Heidegger.

Chris Thornhill

See also Being-in-the-World; Complexity and the Social Sciences; Enlightenment, Critique of; Frankfurt School and Critical Social Theory; Hermeneutics, Phenomenology, and Meaning; Philosophy of Sociology, History of; Spontaneous Order; Structural Functionalism, in Social Theory; Systems Theory

Further Readings

Borch, C. (2011). *Niklas Luhmann.* London, England: Routledge.

Clam, J. (1997). *Droit et société chez Niklas Luhmann: La contingence des norms* [Law and society in Niklas Luhmann: Contingency standards]. Paris, France: Presses Universitaires de France.

King, M., & Thornhill, C. (2003). *Niklas Luhmann's theory of politics and law.* Basingstoke, England: Palgrave.

Thornhill, C. (2012). Luhmann and philosophy. In Niklas Luhmann [Special issue]. *Revue internationale de philosophie, March,* 5–7.

LYING

To a very great extent, social relations and institutions presuppose honest communication. However, communication is not always honest: Individuals and groups deliberately misrepresent reality to others, in the service of personal, political, or ideological ends. In short, people lie. Lying is found at all levels of social intercourse, and its study is therefore of relevance to the entire spectrum of the social sciences. This entry summarizes philosophical conceptions of lying.

At the most general level, we can say that lying is communicative behavior that is undertaken for the purpose of misleading others. Most philosophers accept this characterization, or something very near to it, but they disagree on how it should

be understood. Generally speaking, there are two camps: (1) fine-grained or "narrow" approaches and (2) more coarse-grained or "broad" approaches. Although highly schematic, and therefore inevitably oversimplified, this taxonomy gives a panoramic view of the relevant conceptual landscape.

The dichotomy between the narrow and broad conceptions turns on a difference of opinion about what sorts of behaviors should count as lying. Narrowly conceived, lies are misleading *statements*. A person lies if and only if he or she makes an assertion for the purpose of misleading someone by causing them to embrace a false belief. Although this definition covers a great deal of what we ordinarily mean by the word *lying,* there are some clear counterexamples to it. Consider a boy who cries "Wolf!" with the intention of causing his neighbors to falsely believe that there is a wolf in the vicinity. Even though "Wolf!" is not a statement, we are inclined to think of the boy's behavior as an example of lying. Proponents of the narrow view of lying can respond to this objection by insisting that "Wolf!" is actually a statement in disguise and that in crying "Wolf!" the boy is asserting something like "There is a wolf nearby!" However, there are other cases that are less easily disposed of. Consider groans of pleasure. Groans of pleasure are not statements: They *express* feelings of pleasure without *asserting* them. Now, suppose that a person groans to give a false impression of experiencing pleasure. If the misleading groan is a lie, then the narrow view of lying must give way to a broader one.

According to the broad conception, there are many forms of behavior—both verbal and nonverbal—that are properly included under the umbrella of lying. Removing one's wedding ring to give a false impression of one's marital status, presenting a forged passport to an immigration officer, coloring one's hair to conceal telltale signs of age, and remaining silent in order to hide the truth are just a few examples. Advocates of the broad conception deny that there is no *essential* difference between verbal and nonverbal deception that would justify restricting the term *lying* to the former.

Proponents of both the narrow and the broad approach typically hold that a communicative act is a lie only if it is motivated by an *intention* to deceive. Whether or not the act succeeds is immaterial to its status as a lie. On these accounts, then, there is no such thing as an unintentional lie. Suppose that you are driving along an unfamiliar road and stop to ask for directions to the nearest filling station. Your interlocutor tells you where to go, but unbeknownst to him, the filling station was recently torn down. He misled you, but he did not lie to you, because he misled you inadvertently. Now, contrast the foregoing scenario with one in which your interlocutor deliberately gives you what he believes to be erroneous directions with the intention of leading you astray; however, unbeknownst to him, a filling station had recently been built at the precise location to which he directed you. In this case, the man lied to you about the location of the filling station, even though he did not mislead you. These considerations suggest that it is the *intentions* of the agent, rather than the truth or falsity of his or her assertions, that determine whether or not any given communication is a lie.

There is a third approach to lying—call it the *very broad conception*—which rejects the claim that deceptive intentions are an essential feature of lying. Advocates of this view accept that lying must be purposive, and also accept that intentional behavior is purposive, but they deny that all purposive phenomena are intentional. Biological organs are purposive in function but have nonintentional purposes. For instance, hearts have the purpose of pumping blood around the body, but hearts do not have this purpose by dint of someone's intending it. We can say that biological organs like hearts have *biological purposes* rather than intentional ones.

Nonhuman organisms, even quite simple ones, often mislead one another. For example, the mirror orchid (*Orphys speculum*) produces flowers that mimic the form and scent of female scolid wasps. As a result, male wasps mistake the flowers for females of their own species and attempt to mate with them. In doing this, pollen adheres to their bodies, which they then transport to other flowers. The form and chemical composition of the orchid blossoms have the biological purpose of misleading scolid wasps. That they have this purpose is explained by facts about their evolution. On a very broad conception of lying, then, we can say that mirror orchids lie to the wasps that pollinate them.

The dispute about competing conceptions of lying may appear to be merely semantic. However,

this appearance is misleading, because each conceptual option has ramifications for a variety of fundamental philosophical issues, including the nature of language, testimony, the theory of action, moral psychology, and the relevance of evolutionary biology for explanations in the social sciences.

David Livingstone Smith

See also Moral Cognitivism; Promises and Agreements

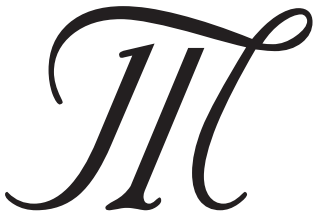
Further Readings

Bok, S. (1978). *Lying: Moral choice in public and private life*. New York, NY: Random House.

Mahon, J. E. (2008). *The definition of lying and deception* (Stanford Encyclopedia of Philosophy). Retrieved from http://plato.stanford.edu/entries/lying-definition/

Martin, C. W. (2009). *The philosophy of deception*. New York, NY: Oxford University Press.

Smith, D. L. (2004). *Why we lie: The evolutionary roots of deception and the unconscious mind*. New York, NY: St. Martin's Press.



Machiavelli's Art of Politics

Niccolò Machiavelli (1469–1527) is without comparison the most accomplished and influential political thinker and social theorist of the Italian Renaissance. Over the centuries, the Florentine's work has given rise to numerous conflicting interpretations and continues to this day to spark debate and controversy. His political writings, state papers, private correspondence, comedies, and poems attract the interest of historians, political theorists, literary scholars, and students of classical republicanism alike. Outside academia, politicians, political commentators, and public figures continue to draw on his work as a manual for political success and a source of worldly wisdom. The strategic aspects of his writings and his penetrating analysis of human motivations and power relations have, in a burgeoning literature on leadership and management, come to be applied to the world of corporate business and enterprise. Machiavelli's status as a major classical thinker cannot be denied.

Machiavelli's originality as a political thinker and as a social theorist is intimately linked to his method. As the opening statements in three of his main works, *The Prince* (1513), the *Discourses on Livy* (ca. 1514–1518), and *The Art of War* (1520), make clear, his approach to politics is based on a dual optics. On the one hand, he shares the Renaissance humanist's penchant for ancient history and literature, contending that human nature and the basic conditions of political and social life have not undergone any major changes since antiquity. Consequently, lessons from ancient history, if interpreted correctly, can be applied to modern conditions, offering principles, maxims, and general rules of conduct upon which a well-balanced and vigorous republic can be erected. On the other hand, Machiavelli inaugurates a new type of political theorizing based on first-hand observations of the political affairs of the day. Here, his personal experiences as a diplomatic envoy and as a military observer come to inform a general outlook that places great emphasis on the exigencies of the political here-and-now. In Machiavelli's work, these two perspectives superimpose and blend into a new type of pragmatic and classically inspired political discourse.

Machiavelli's art of politics is centered on his innovative notion of virtue. While traditional Aristotelian ethics had defined virtue as a mean between two extremes, Machiavellian virtue, in contrast, involves a capacity to employ the extremes according to the circumstances. The Machiavellian prince or statesman should thus abide by the law and conventional morality when he serves his or his republic's interest, but he must be prepared to use deception and force, metaphorically described as the fox and the lion, when so required.

The underlying logic of Machiavelli's notion of virtue is illustrated by his semiconceptual notion "cruelties well used" (*crudeltà bene usate*). The violent methods employed by the ruthless Cesare Borgia qualify as "well used," since they enabled the ruler to lay a strong foundation for his power, which later allowed him to adopt a more civil and peaceful form of rule. As the example of Cesare Borgia teaches, the means employed are retroactively justified by the political outcome. By arguing that cruelty

(and other traditional vices) can be a political, and perhaps even a moral, good, Machiavelli dislodges the very notions of good and evil from their theological underpinnings. In Machiavelli's art of politics, cruelty and, consequently, the other traditional vices are no longer regarded as an absolute evil but are conceived of as instruments of rule that can be used well or badly according to the circumstances and the designated ends.

One of Machiavelli's main contributions to social theory consists in his conception of social conflict and partisan strife as fundamental facts that legislators and rulers need to acknowledge, contain, and channel. In keeping with the main tendency of Florentine republicanism, Machiavelli argues that a republic has a dual aim: one internal, associated with the classical concept of liberty, and the other external, aspiring to acquisition of dominion (*imperium*), material goods, greatness, and glory. In the republicanism of the *Discourse*, an intimate link exists between this dual pursuit of liberty and empire, on the one hand, and the republic's mixed constitution and internal ordering, on the other. According to Machiavelli, there are, broadly speaking, two categories of men: (1) those who desire not to be oppressed but to live as free individuals and (2) those who want to command and dominate others. The former group Machiavelli calls the people (*popolo*) and the latter the great (*grandi*). From the viewpoint of classical constitutionalism, the two categories correspond to the many and the few, respectively. One of the chief concerns of the Machiavellian statesman is to find artful ways of satisfying, or accommodating, each of these contrasting desires, or appetites, without suppressing the demands of the other, conflicting humor, or category, and without endangering the republic as a whole.

As a true partisan of the mixed regime, Machiavelli refuses to side with either of the two categories, and therefore, his republicanism cannot easily be characterized as either popular or elitist. At the same time as he praises the ancient Roman elite's manipulation of the populace's religious beliefs for political ends, and exhorts his contemporaries to revive it, as he puts it, by reinterpreting Christianity "according to virtue," he endorses an aggressive and ferocious form of popular republicanism in which the general populace, acting as "the guardians of liberty," controls the elites by often harsh and brutal methods.

By taking human beings as they are, with all their imperfections, partialities, and vices, as his starting point, rather than humankind as it ought to be, Machiavelli occupies a central position in the prudential and realist tradition in Western political philosophy. For him, as for thinkers like Thucydides, Aristotle, Montesquieu, Madison, Burke, and Tocqueville, politics is not a means for implementing a philosophical ideal but an inherently messy business, full of contradictions and conflicts. In its Machiavellian variety, this pragmatic approach involves developing a double viewpoint that comprises the irreconcilable demands of liberty and empire. Consequently, the Machiavellian art of politics consists of balancing the interests of the modest and peace-loving citizen and the exceptional individual's quest for greatness, conquest, and glory. Insofar, it can be argued, it addresses basic aspects of the human condition and comes to acquire a timeless quality.

Mikael Hörnqvist

See also Hobbes's Philosophical Method: Nature–Man–Society; Philosophy of Politics, History of; Vico's *Scienza Nuova*

Further Readings

Hörnqvist, M. (2004). *Machiavelli and empire*. Cambridge, England: Cambridge University Press.

Machiavelli, N. (1988). *The prince* (Q. Skinner & R. Price, Eds.). Cambridge, England: Cambridge University Press.

Machiavelli, N. (1997). *Discourses on Livy*. Oxford, England: Oxford University Press.

Mansfield, H. C. (1996). *Machiavelli's virtue*. Chicago, IL: University of Chicago Press.

Skinner, Q. (1978). *The foundations of modern political thought: Vol. 1. The Renaissance*. Cambridge, England: Cambridge University Press.

Skinner, Q. (1981). *Machiavelli*. Oxford, England: Oxford University Press.

Stacey, P. (2007). *Roman monarchy and the renaissance prince*. Cambridge, England: Cambridge University Press.

Viroli, M. (1998). *Machiavelli*. Oxford, England: Oxford University Press.

Machine Consciousness and Autonomous Agents

Machine consciousness (MC) concerns the development of computer models of human mental features that are closely associated with consciousness

in humans and other biological creatures. How far can consciousness be modeled or replicated in an inorganic machine? The study of MC is closely associated with the larger and more established field of artificial (or machine) intelligence (AI). MC is sometimes referred to as "artificial consciousness," but the latter is arguably wider in scope. For instance, one could imagine creating complete conscious biological beings by building and growing tissues in a lab using possible future biosynthesis techniques; they would perhaps count as being artificially conscious but not as MCs. The specific challenge of MC is to see how far developments using the well-understood computer-based technologies available today can shed light on consciousness.

MC and Artificial Intelligence

Clearly, there is a close association between AI and MC and, indeed, between cognition and consciousness. Humans and many animals need to be consciously aware to exercise various kinds of intelligence or cognition. Tasks such as checking for predators or dangers or concluding a chess endgame are usually best performed in a condition of concentrated alertness. However, other kinds of cognition—visual processing, for example—seem to occur, to a substantial degree, below the threshold of consciousness. There is much debate over what kinds of cognition are integral to consciousness: Many argue that the "raw feels" of sentience, which are possibly present even in species with relatively primitive nervous systems, are quite distinct from cognitive processing. Nevertheless, "higher" forms of consciousness, such as self-awareness or self-monitoring, are probably integrally linked with cognitive processing. Ironically, but perhaps not too surprisingly, MC researchers seem to have more success in modeling more sophisticated kinds of consciousness than the more primitive types; this encourages sceptics about MC to say that such models are really providing insight only into the cognitive *accompaniments* to consciousness, rather than into states of consciousness in themselves.

"Strong" and "Weak'" MC

It has been customary for decades to distinguish between "weak" and "strong" AI. Supporters of strong AI believe that computers will one day literally have mental states (and perhaps, they already do). Supporters of weak AI do not accept this but still maintain that computer models can give important insights into the nature of intelligence and mind. A similar distinction can be made between weak and strong MC. (It is probably best to talk about a weak–strong spectrum in each case.) For weak MC, computational or robotic techniques provide useful models of consciousness but not real consciousness in machines, whereas strong MC suggests that such machines are possible and may even be imminent. Strong MC shares with strong AI an acceptance of the philosophical functionalist view, which asserts that the mind is best explained in terms of abstract computational operations of the brain, or of functionally equivalent hardware, and which emphasizes in particular that consciousness can be explained without residue in cognitive/computational terms. Weak MC dissents from such strong claims about the theoretical relation between mind and computational processes.

Current work on MC follows a number of hypotheses about cognitive processes that may be necessary or sufficient for a system to be conscious. Several researchers work on developing computational or robotic systems that implement existing theoretical frameworks, such as Global Workspace theory, or self-modeling approaches to consciousness. Other approaches seek to build artifacts that instantiate specific cognitive faculties, such as imagination, that may be sufficient for consciousness, or a broad set of cognitive functions that may be jointly sufficient. Some others set out to challenge common understandings of consciousness—for example, by building models that depart from the idea of a singular "stream of consciousness." Most current work does not explicitly attempt the (strong MC) target of building actual conscious machines but rather holds to the (weak MC) idea that building models of various processes underlying consciousness will produce new insights about consciousness as such. However, successful work of the latter kind may well reinforce confidence in future work directed at the strong-MC target.

Ethical Aspects of MC

Weak MC raises many interesting technical issues, but strong MC (like its AI counterpart) raises some important challenges concerning how far computers and robots can possess a mental life. The quest to create computer-based consciousness arguably introduces a strong ethical element into the debate. Mainstream AI and robotics raise ethical issues to

do with the responsibilities of those who design and deploy such cognitive agents. Computer-based agents that were conscious would also raise questions to do with our responsibilities toward such agents themselves. In the late 18th century, the English philosopher and advocate of social reforms Jeremy Bentham (the father of utilitarianism) argued that nonhuman animals deserve our moral consideration, not because they reason but because they *suffer*, that is, have sentient awareness of harm done to them. One supposes that Bentham would have wished to extend his ethical concern to potential artificial agents if they, too, had conscious awareness of their states.

The link between conscious awareness and moral concern seems to be difficult to deny. So if some machines were admitted to be genuinely conscious, in the sense of being capable of suffering, feeling pain or fear, and so on, then arguably they should be deemed to have genuine moral interests. As with animals, their inclusion in our moral constituency would be amplified if they had higher, more cognitively scaffolded forms of consciousness. If such MCs were to proliferate, this would surely create some deep social conundrums for future generations, particularly if the conscious mental life of such beings were to be enhanced by levels of intelligence far greater than those of humans, as seems highly possible given the current levels of attainment in some branches of AI.

Strong MC is perhaps ethically or socially significant for another kind of reason. Artificial systems that incorporated higher, more self-reflective, forms of consciousness could possibly be considered as moral or social subjects, not just in the "receptive" sense just mentioned but also in the sense of being *autonomous agents* that could deliberate over and be given moral credit or blame for their actions and choices. Being a morally responsible agent, arguably, requires being capable of understanding what it is like to experience the positive or negative effects of different courses of action on the course of experience of the people affected. Such an understanding may in turn require being capable of having such experiences oneself, at least in principle.

Some researchers in the field of machine ethics are developing ethical deliberation systems that arguably have some key features of conscious agents. For example, they claim, such systems may be considered morally autonomous agents to the extent to which their mental organization involves a kind of global integration that many people believe to be crucial to self-awareness or consciousness. So the enterprise of modeling of consciousness and of modeling ethical thinking may be closely associated.

Scientific Versus Engineering Motives for MC

Why do work in MC? One can distinguish between scientific and engineering—or theoretical and practical—motives for doing research in MC. (Again, a similar distinction can be made within AI research.) Someone pursuing scientific MC would be primarily motivated by the wish to understand the nature of consciousness and to explain it in scientific, or maybe philosophical, terms. Someone pursuing engineering MC would be rather motivated by the idea that if we can replicate or simulate consciousness in artificial systems, they are likely to work better for us. As mentioned above, we generally do things better when we consciously attend to them: Our consciousness allows us to be more sensitive or responsive to our environment, and so on.

Of course, a given research program may be fueled by both motives, but they may pull in different ways. It would be disingenuous to criticize an engineering design project that seeks to incorporate characteristics of self-awareness into a plant control or aircraft navigation system simply on the grounds that it gives little insight into the nature of consciousness per se.

MC, Attached and Unattached

A further distinction worth drawing is between attached and unattached MC. *Attached* (or "prosthetic") MC concerns the development of devices that may provide an already conscious individual with enhanced or additional kinds of consciousness. *Unattached* (or "autonomous") MC is targeted at developing self-standing systems that have consciousness in their own right. Seemingly, the latter provides a deeper challenge, since the former seems to "piggyback" on an existing consciousness. It is relatively easy (while still deeply challenging, of course) to integrate devices that modify an already existing conscious creature, as compared with progressing from sophisticated but phenomenally null computational systems to computational agents that have their own independent experiential awareness.

There have already been a number of breakthroughs in neural implant or perceptual supplementation technologies that count as forms of attached or prosthetic MC. An example is the tactile–visual

sensory systems developed by Bach-y-Rita and others, where blind subjects are equipped with arrays of tactile stimulus points on the back or tongue, creating patterns of two-dimensional stimulation corresponding to input from a camera attached to the head. Fluent users of such devices report experiences that are vision-like (e.g., including depth perception) but not truly visual, while not tactile either. Such devices are important in expanding our conception of what might be involved in consciousness and therefore provide useful support for the idea that computer-powered agents may have forms of experience or phenomenology that are quite different from human phenomenology but that still deserve to be called consciousness.

Human MC Versus General MC

Human consciousness is what we know best. Some would define consciousness in such a way that only humans, or beings very much like them, are conscious: for example, by requiring of conscious creatures that they be capable of certain forms of higher-order thoughts or perceptions or be otherwise self-reflective in ways that require a sophisticated cognitive apparatus involving, in the natural world, human or primate levels of intelligence. Such conceptions of consciousness, of course, lend themselves to various kinds of AI or robotic developments that mimic rich human cognitive and bodily capabilities. However, one approach to MC research seeks to replicate the consciousness of more primitive creatures—for example, concentrating on fairly simple embodied or "enactive" aspects. Other avenues of MC research deliberately avoid slavishly copying the biological realm at all, targeting only the most abstract features of existing conscious organisms in order to stretch the envelope of possibilities for what it is to be a conscious agent or system.

Super-Consciousness in Super-Intelligent Machines?

Another way in which one can extend the possibilities for MC—at least speculatively, if not currently, in terms of working systems—is by seeing human-level forms of consciousness and intelligence as a baseline from which one may go up as well as down. Thus, some researchers inspired by transhumanism, for example, may favor the view that as systems develop progressively more complex and expanded kinds of cognitive attainment, rich forms of consciousness—perhaps "superconsciousness"?—will be present in those systems, with enhanced kinds of experiences that are in some sense deeper or "higher" than those available to human subjectivity. Some are even envisaging a situation where the informational contents of individual human brains may be uploaded onto a silicon base, allowing one a form of personal, experiential survival—including a subjective sense of psychological continuity with one's former self—after the decay of one's original body. Talk of such possibilities raises deep issues concerning the philosophy of personal identity, as well as further profound questions concerning the way technologies may shape our sense of what it is to be a social being and to be conscious.

MC and Autonomy

There are many other kinds of human-like properties, besides consciousness and intelligence, that are often discussed in relation to machines. In particular, people often talk of machine "autonomy"—for example, agents that can think, plan, act, and so on, "for themselves" rather than only under direction. There are a number of ways in which the term *autonomy* can be taken, which are relevant to investigations in MC.

In its most basic sense, an *autonomous agent* is one that can choose, plan, and act "for itself" rather than merely act under external control or guidance. An airborne attack drone could be autonomous in this sense (most are remotely controlled by humans at present, but this could change). One might call this *operational* autonomy. An agent that is operationally autonomous in this minimal way does not seem to be a candidate for being considered conscious or self-aware in any deep sense.

Operational autonomy can be distinguished from *normative* autonomy. The latter might apply to a robot that is considered to be independently responsible or accountable for its decisions and actions. It is not clear what might be necessary for autonomy in this latter, stronger, sense. Arguably, some kind of conscious awareness might be a prerequisite, but this could be challenged.

A third, but related, kind of autonomy would be *social* autonomy: An agent might be considered to be a member of society in the sense of having particular kinds of social or legal rights and/or obligations. For example, social autonomy might entail rights and duties that go with the ownership of property or

wealth or with having other kinds of social status. Some kind of conscious awareness may be thought to be necessary for autonomy of this latter kind, but again, this can be disputed. In many jurisdictions, commercial corporations are considered as legally autonomous "persons," which can own property, and so on, without their being deemed to have any independent consciousness (over and above the humans who co-own or operate them). So it is a moot point if attributing social autonomy to an artificial agent implies the attribution of consciousness.

A final kind of autonomy worth mentioning is *biological* autonomy. It is often claimed that any self-standing biological organism possesses a kind of autonomy (sometimes referred to as *autopoiesis*) that enables that organism to differentiate itself from its surroundings and to maintain itself, via a set of chemical and metabolic processes, as a continuing living system. It has often been asserted that no human-engineered machine—at least if built using existing technologies—could be autonomous or autopoietic in this deep sense and, further, that it is just this lack of biological autonomy that makes the development of a genuine MC impossible or even paradoxical.

This is also a contested claim; but if it were true, then MC would be a far more limited field than many of its enthusiasts would accept. However, perhaps some form of artificial autopoiesis could be developed using computer-modeling techniques. If so, perhaps a machine that had something akin to a biological autonomy, but without itself being a biological system, could be perfected. It is far from clear what would be involved in such a possibility, but investigating it both theoretically and experimentally might be an important future development path for the field of MC, as well as for the understanding of consciousness in its natural, biological forms.

Steve Torrance and Robert Clowes

See also Agency; Agent-Based Modeling and Simulation in the Social Sciences; Artificial Intelligence; Classical Computationalism, Connectionism, and Computational Neuroscience; Cognitive Sciences; Consciousness; Human–Machine Interaction

Further Readings

Aleksander, I. (2005). *The world in my mind, my mind in the world*. Thorverton, England: Imprint Academic.

Blackmore, S. (2010). Artificial consciousness. In S. Blackmore (Ed.), *Consciousness: An introduction* (2nd ed., sec. 6). London, England: Hodder Education.

Chalmers, D. (2010). The singularity: A philosophical analysis. *Journal of Consciousness Studies, 17*(9–10), 7–65.

Chrisley, R. (2008). Philosophical foundations of artificial consciousness. *Artificial Intelligence in Medicine, 44*(2), 119–137.

Clowes, R. W., & Seth, A. (2008). Axioms, properties and criteria: Roles for synthesis in the science of consciousness. *Artificial Intelligence in Medicine, 44*(2), 91–104.

Gamez, D. (2008). Progress in machine consciousness. *Consciousness and Cognition, 17*(3), 887–910.

Holland, O. (2003). Editorial introduction. In *Machine consciousness* [Special issue]. *Journal of Consciousness Studies, 10*(4–5), 1–6.

Holland, O. (Ed.). (2003). *Machine consciousness*. Thorverton, England: Imprint Academic. (Also published as a special issue of *Journal of Consciousness Studies, 10*(4–5))

Shanahan, M. (2010). *Embodiment and the inner life: Cognition and consciousness in the space of possible minds*. Oxford, England: Oxford University Press.

Torrance, S. (2007). Two conceptions of machine phenomenality. In *Machine consciousness: Embodiment and imagination* [Special issue]. *Journal of Consciousness Studies, 14*(7), 154–166.

Torrance, S., Clowes, R., & Chrisley, R. (Eds.). (2007). *Machine consciousness: Embodiment and imagination* [Special issue]. *Journal of Consciousness Studies, 14*(7).

Markets and Economic Theory

This entry introduces the basic aspects of the institution of the market, describing its normative significance and basic functions. It then proceeds to examine the controversial relationship of markets to the distribution and consumption of public goods and concludes by raising some critical points.

Foundational Aspects

Economists see markets as the primary institutional site for bilateral exchange. Market exchange is *normatively* significant because the parties to the exchange are both made better off in expected terms and because exchange permits the separation of production and consumption activities, which facilitates specialization and the division of labor. The first aspect allows individuals to benefit from natural differences in tastes and talents and (in the case of international trade) differences in climate and

institutional arrangements. The second aspect allows individuals to take advantage of the productivity increases that specialization creates—increases that, according to Adam Smith, are the primary source of human progress. In this sense, markets are first and foremost an institutional device for mobilizing the benefits of human cooperation. Cooperation rather than competition is the central feature of the market order.

Basic Functions of Markets

Because market exchange is essentially *voluntary*, within the bounds set by an initial assignment of the rights and rules of contract (e.g., rules against fraud), markets are also seen by some philosophers as instantiations of *free* interactions—though most economists value "liberty," so understood for its consequences for material well-being rather than intrinsically.

A well-functioning free market performs *three functions*: (1) it allocates the available supply of goods among individual demanders according to the intensity of their effective demands (their willingness to pay), (2) it ensures that aggregate supply is distributed among suppliers so that demanders are supplied at minimal cost, and (3) it connects demand to supply by providing both the information and the incentive that suppliers require in order to meet the prevailing demand.

Nonmarket mechanisms for the allocation of goods (such as queuing or rationing) will routinely fail in one or other of the three aforementioned functions. Queuing, for example, may allow demanders to express their demand (by waiting in line) and allow suppliers to assess demand (by examining the length of the queue). But in nonmarket institutions, the waiting time given up by queuers does not translate into any benefit to suppliers, so suppliers have no incentive to adjust supply in light of the demand conditions: The critical link between demand and supply that a market would provide is thus severed.

The informational and incentive features of markets are worth special (and separate) emphasis. Within the free market system, prices constitute signals as to the value that others place on alternative activities that potential suppliers might engage in. Prices operate as a kind of gauge for measuring the effect that my choices have on the well-being of others (as they themselves assess it). In the absence of market prices, such information would be unavailable (a point mobilized against those who might look to socialist planning to simulate the beneficial effects of markets). Furthermore, in the market system, each player has an incentive to act in the way others "most prefer": The beneficial effects of markets emerge "invisibly"—that is, without those beneficial effects being the central intention of any of the participants. Put another way, the benefits of human cooperation are reaped with only a minimal claim on the players' dispositions to act "cooperatively." Achieving the beneficial effects of markets does require that participants refrain from fraud and deception—but, to some extent, markets themselves provide rewards for individuals who so refrain. A reputation for reliably good service is an asset in market transactions, especially for products whose quality is not easy to assess on inspection.

Markets and Public Goods

Although markets work broadly in the public interest for ordinary "private" goods, like apples and houses, they often fail to work well in the case of so-called public goods. Public goods are goods that all members of a relevant community consume in common, in such a way that none can be excluded from the benefits. More specifically, public goods have two properties: (1) they are *non-rival*, in the sense that each consumer enjoys full use of the good without diminishing the amount available for the other users (as in the case of a concert performance, where an "encore" for one consumer means an encore for all the others), and (2) they are *non-excludable*, in the sense that individuals who do not pay cannot be prevented from consuming what is available. Both non-rivalry and non-excludability can come in degrees, and either can be present without the other. Economists often associate public goods with "market failure," since rational actors in a market will fail to produce goods for which there is potential demand when they lack the ability to charge consumers.

Market failure is typically identified as providing scope for government intervention in markets. To be sure, many of the standard activities of governments—the provision of a legal system, with enforcement of rights; provision of national defense; and certain kinds of public health provision (including pollution control and vaccination programs)—fit the public goods categorization. But many others (such as private health provision and public housing subsidies) appear to exemplify the public goods category much less well. In these latter cases, the

standard results about market success and market failure do not support government intervention.

But nor do those results in themselves argue *against* government intervention. Any case for or against such intervention (in relation to public or private goods) requires a corresponding examination of the normative properties of the relevant institutional alternative. Such "examination" involves a specification of the incentives of agents operating in democratic political settings and an investigation of the ways in which "public supply" can be expected to respond to "citizen demand." Undertaking this exercise has been the motivation for the development of "rational actor political theory" (or "public choice theory").

Critical Issues

The general point here is that no discussion of the normative properties of any institutional arrangement—including markets—can be undertaken without an answer to the "compared with what?" question. Claims about market success and market failure only have normative bite when the analogous analysis of alternatives (primarily government action) has been completed.

Markets are often criticized—even by those who accept their many advantages—on the grounds that they give rise to distributions of income that are "unjust" or, more modestly, "could be made more equal." We do not here explore this line of criticism. We note, however, that even here the normative exercise is ultimately a comparative one: If the market fails to provide a distribution of income that meets all normative standards, it remains to be shown that the distributional modifications that emerge under alternative political institutions will predictably improve things.

Geoffrey Brennan and Jonny Anomaly

See also Austrian Economics; Capitalism; Homo Economicus; Libertarianism, Political; Pareto Optimality; Public Goods; Sen's Paretian Liberal

Further Readings

Anomaly, J. (2011). Public health and public goods. *Public Health Ethics, 4*(3), 251–259.

Brennan, G., & Eusepi, G. (Eds.). (2009). *The economics of ethics and the ethics of economics: Values, markets and the state*. Cheltenham, England: Edward Elgar.

Brennan, G., & Lomasky, L. (1997). *Democracy and decision: The pure theory of electoral preference*. Cambridge, England: Cambridge University Press.

Graafland, J. (2007). *Economics, ethics and the market: Introduction and applications*. London, England: Routledge.

Hayek, F. (1945). The use of knowledge in society. *American Economic Review, 35*(4), 519–530.

Miller, D. (1989). *Market, state, and community: Theoretical foundations of market socialism*. Oxford, England: Clarendon Press.

Smith, A. (2010). *The wealth of nations*. New York, NY: Modern Library. (Original work published 1776)

Marxism and Social/Historical Explanation

This entry reviews Marx's theory of social and historical explanation and his own methodological stance vis-à-vis what counts as "science," explains what is distinctive about it, shows the way the theory was subsequently developed or altered, and ends by introducing one of the most formidable critiques and an equally powerful defense of it in the 20th century.

Introduction

Explanation as a distinctive methodological concept and problem was posed during the development of the social sciences in the later 19th and early 20th centuries. It, thus, postdates the original practices of history writing by more than 2,000 years, though various historians, beginning with Thucydides in ancient Greece, have commented on good and bad ways of being a historian. On the other hand, social science, including social theory, is a modern intellectual and academic set of disciplines, which has self-consciously generated and absorbed various methods and formalized methodologies, in particular those derived from or projected onto the physical sciences.

The methodological problems inherent in social/historical explanation thus largely postdate the major works of Karl Marx (1818–1883), though not the development of the various Marxisms through which his thought was later interpreted. Within those Marxisms, there have been notable claims and consequent debates through which the current methodological pluralisms of the social sciences and

historiography have developed. The most famous of these controversies was in the later 20th century and concerned the concept of explanation in relation to "the materialist conception of history."

Marx and "Science"

Marx considered his own work to be *wissenschaftlich*, usually translated as "scientific" but better understood as "systematic" and "rigorous." As construed by preceding German philosophers, this outlook did not pose a dichotomy between philosophy and other humanistic studies, on the one hand, and the physical sciences, on the other. Rather, it was presumed that all studies were amenable to *wissenschaftlich* treatment and methods of inquiry. This reflects the Aristotelian and Thomistic ambitions and concepts of "school philosophy" and also profoundly rejects early-modern Anglophone empiricisms that privileged a materialism of sensory experience. Thus, it is *a mistake to project a dichotomy between the natural or physical sciences and the social or human sciences onto Marx's own theorizations of history and society*. Similarly, it is also a mistake to presume that he accepted a hierarchy of methods in which the physical sciences ranked more highly than historical or other humanistic forms of knowledge production.

Marx's most original work in social theory is his claim that *historical progress (or otherwise) is marked by changes in economic practices and structures*, in particular changing technologies of production and varying systems of property and other legal relations. He occasionally identified his studies as "empirical" and his presuppositions as "material," but his works have been commonly misconstrued as endorsing various philosophical materialisms, often said to be coincident with those of the physical sciences. In his writings, Marx conceptualized human social systems *historically*, arguing that productive activities were necessarily pursued there, but always within constitutive conceptual frameworks that have distinctive histories and logical structures. Common to these historical formations were relations of domination and oppression, which he termed *class struggle*.

Marx was critical of synchronic and individualistic social theory, charging famous political economists of preceding generations with ignoring history and effectively promoting the interests of property owners, whether in land, trade, or manufacture. He also criticized conventional materialisms, even when construed as coincident with the physical sciences, and rejected the presumptions of the political economists and, by extension, of modern economists in the marginalist, mathematical tradition.

Marx was thus not particularly concerned to formulate protocols of "explanation" or to adhere closely to any particular vocabulary that would make his writings "scientific." His self-conscious comments on his work and methods were not extensive, and his now most famous effort, the Preface to *A Contribution to the Critique of Political Economy*, published in 1859 (but not widely circulated until 1897), employs a bewildering array of important concepts in relations of ambiguous synonymy. It uses the verb "explain" (*erklären*) but once and then in an apposition with the verb "judge" (*beurteilen*).

Engels and the "Materialist Dialectic"

This was not the case with Marx's close friend, political associate, and occasional coauthor Friedrich Engels (1820–1895). In 1859, he began to popularize Marx's first published version of his lifetime project, namely, a critique of political economy. Engels broached the issue of method, attributing a unique set to Marx. He parsed Marx's methodology into logical and historical divisions but failed to sustain these distinctions. However, returning to the task of conveying Marx's thought in summary form in the later 1870s, Engels published *Herr Eugen Dühring's Revolution in Science* (commonly known as *Anti-Dühring*). This was ironically titled with reference to revolution but not to science. In the intervening years, Engels had adopted an outlook that privileged the physical sciences in terms of a matter-in-motion materialism. But he also claimed to recoup what he took to be the Hegelian dialectic of progress through conceptual contradiction and developmental resolution.

Asserting that the two sides of this dichotomy were in fact coincident, Engels argued that this was the foundation of Marx's work and the guarantor of its truth in relation to both historical explanation and political prediction. In this narrative, he mirrored the intellectual currents later generalized as positivist or empirical philosophies of scientific explanation. This was a view that all sciences establish facts through specific methodologies determined by logic and nature. Explanation occurs when facts that demonstrate regularities are assembled under

laws that possess predictive power. By publishing a formidable set of new prefaces and introductions to new editions of Marx's works between 1883 and 1895, Engels effectively inserted Marx into the frameworks through which "explanation" was arising as a problem in social science. He also outlined a solution to which many Marxists adhered but against which not a few rebelled.

Historical and Dialectical Materialisms

At the close of the 19th century and during the first decade of the 20th, G. V. Plekhanov (1857–1918) and Karl Kautsky (1854–1938) circulated influential works on what Engels had formulated in 1859 as the materialist conception of history. These reproduced Engels's later methodological claims of the 1870s that Marxism comprises a universal science based on laws—common to nature, history, and thought—that would be explanatory and therefore predictive. The predominant assumption in this orthodoxy was that causal and predictive determinism—modeled on the supposed certainties of the physical sciences—would give Marxists a real advantage over their other socialist rivals.

However, both thinkers, and those in their respective Marxist communities in Russia and Germany, also grappled with the specifics of explaining particular historical and contemporary social formations. In political terms, they also addressed the apparently intractable problem of individual agency, given that "history" and "thought" were conceptualized as effects of processes that were more fundamental, more regular, more effectual, and more predictable than the apparent "voluntarism" of human decision making. Both thinkers therefore confronted contrary views as to how correct and efficacious explanations for social phenomena should be constructed; Plekhanov attempted to integrate a view of individual action into the explanatory architecture of historical materialism without contradiction, and Kautsky defended a deterministic view against the apparent contradiction generated by introducing an ethical element into Marxism.

Max Weber's (1864–1920) early-20th-century essays on "The Protestant Ethic and the Spirit of Capitalism" represented a challenge both to Marx's emphasis on the development (or otherwise) of productive activities as crucially important in historical change and in structuring social life and to Marxist claims that lawlike regularities—understood as ultimately determining for human development—were the most valid explanatory model for understanding society and politics. The distinction Weber invoked between *erklären* (clarification with objectivity) and *verstehen* (understanding with subjectivity) effectively set up the problem of explanation for the social sciences during the rest of the 20th century.

Acton's Attack and Cohen's Defence

In 1955, the British philosopher H. B. Acton (1908–1974) published a full-length polemical attack on Marxism, which he took to be a worldview based on dialectical and historical materialism. He argued that Marxism was incoherent because it attempted to combine positivism—by accepting modern science and its methods—with metaphysics—in particular Hegelian or "speculative" notions of development through contradiction. Marxists were thus projecting contradictions (which could not be empirically observed) into material phenomena (which could), so Marxism was consequently a farrago of idealism and materialism.

In his longest chapter, Acton argued that the materialist interpretation of history was either anodyne (economic factors are always of some interest and influence in historical explanations) or formulated in terms of untenable or incoherent distinctions (e.g., "basis/superstructure" or "productive forces" vs. "productive relationships"). He concluded that the Marxist version of technological determinism could not possibly be true, because it was untestable against facts and was therefore, in his view, unscientific. It was "the illusion of the epoch."

In the later 1960s, the Canadian philosopher G. A. Cohen (1941–2009) began work in the United Kingdom on a refutation of Acton's critique, though in philosophical terms quite similar to Acton's and in terms of Acton's own formulation of the issues. Eventually published in 1978, *Karl Marx's Theory of History: A Defence* caused a considerable stir in Anglophone philosophical and social science circles by claiming that if clearly and parsimoniously reformulated and if rigorously tested against historical evidence, the materialist interpretation of history as espoused by Marxists could be salvaged and successfully defended.

Following Acton, Cohen understood historical materialism as a theory of explanation applicable

to social structures and sociopolitical change. He argued that the level of development of productive forces or technologies had explanatory primacy in human history because humans were somewhat rational, possessed intelligence, and faced scarce material resources. However, Marxists had construed their explanatory claims as scientific in form and resting on fact, whereas Cohen contrarily construed these claims—reformulated by him in propositional statements—as "functional" explanations, of which Marxists and many others had always been suspicious.

In Cohen's account, functional explanations demonstrate that the character of a thing to be explained is determined by its effect on the thing that explains it. For example, production relations have a profound effect on productive forces, yet the character of productive relations is explained by the level of development of the productive forces themselves. The chain of causation thus runs opposite to the directionality of the explanation; that is, x has an effect on y, yet the character of x is explained by its functional (rather than causal) relationship with y.

Merely identifying various existent or possible functionalities did not in itself generate explanations. Cohen argued that further empirical elaborations were necessary and that, when rigorously formulated and satisfactorily evidenced, such functional explanations could be true. In Cohen's view, Darwin's theory of the origin of species by natural selection, when convincingly elaborated in specific cases, shows in what manner the useful equipment possessed by plants and animals accounts for the existence of these observable features and utilities in plants and animals themselves.

Cohen's chapter on functional explanation in the social sciences attracted even more interest than his defence of historical materialism, not the least because it addressed the social sciences and philosophies of science more generally. After more than a decade of debate, however, he withdrew his defense of functional explanation and also his defense of the materialist interpretation of history, saying that even in his own rigorous reformulation it was unfortunately false.

Displacement of the Problem

The terms of the problem of explanation were set by presumptions that the physical sciences represent a methodological ideal for the social sciences. Thomas Kuhn's (1922–1996) *The Structure of Scientific Revolutions* (1962) undermined this idealization of science and its supposed methodological unities, though Cohen chose not to engage with this challenging work. The development of poststructuralism, and in particular the "linguistic turn" in discourse analysis, together with the influence of Foucauldian archaeologies and genealogies, has for many historians and social scientists displaced methodological issues having to do with explanation and science. Marx's general ideas survive, loosely construed, in the construction of social and historical explanations, but the impetus to update Marxism as a specifically scientific and/or explanatory theory has passed.

Terrell Carver

See also Capitalism; Determinism; Dialectic, in the Social Sciences; Empiricism; Explanation, Theories of; Explanation Versus Understanding; Holism, in the Social Sciences; Marxist Economics; Positivism, History of

Further Readings

Aston, T. H., & Philpin, C. H. E. (Eds.). (1987). *The Brenner debate: Agrarian class structure and economic development in pre-industrial Europe*. Cambridge, England: Cambridge University Press.

Callinicos, A. (2006). G. A. Cohen and the critique of political economy. *Science & Society, 70,* 252–274.

Carling, A. (2006). Karl Marx's theory of history and the recovery of the Marxist tradition. *Science & Society, 70,* 275–297.

Carver, T. (1982). *Marx's social theory*. Oxford, England: Oxford University Press.

Cohen, G. A. (1978). *Karl Marx's theory of history: A defence*. Princeton, NJ: Princeton University Press.

Farr, J. (1984). Marx and positivism. In T. Ball & J. Farr (Eds.), *After Marx* (pp. 217–234). Cambridge, England: Cambridge University Press.

Nolan, P. (2006). Why G. A. Cohen can't appeal to Charles Darwin to help him defend Karl Marx (but why others can). *Science & Society, 70,* 155–179.

Rigby, S. (1998). *Marxism and history: A critical introduction* (2nd ed.). Manchester, England: Manchester University Press.

Wood, E. M., & Foster, J. B. (2006). *In defence of history: Marxism and the postmodern agenda*. Delhi, India: Aakar Books.

Marxist Economics

Neither Karl Marx nor Friedrich Engels left a detailed explanation of the philosophical underpinnings of Marxist economics. Therefore, Marxist economists and other scholars of Marx's critique of political economy have had to interpret the philosophy at work in Marx's actual texts on economics, especially *Capital*, in order to understand and further develop Marxist economics. This entry includes an introduction to Marx's main texts on his philosophical approach to economics as well as an explanation of the three major interpretations of that philosophy: modernism, postmodernism, and critique.

Marx

The key source of Marxist economics is Marx's *Capital*, in three volumes. In that text, Marx develops his "critique of political economy" (the subtitle of *Capital*), in two senses: (1) a critique of capitalism and (2) a critique of the mainstream economic theory of his day that celebrated capitalism. Marx criticized capitalism because it is based on the exploitation of workers by capitalists and because it leads to recurring economic crises, and he criticized the classical political economists (e.g., Adam Smith, David Ricardo, and others) because they ignored exploitation and failed to see that capitalism's own dynamic creates periodic crises.

Marx (and, after he died, Engels) offers a few words of explanation of the philosophy guiding his writings on economics in the various prefaces and postfaces to *Capital*. But scholars tend to look to other writings, especially the Preface to *A Contribution to the Critique of Political Economy* and the Introduction to the *Grundrisse* (the notebooks Marx wrote in preparation for writing *Capital*), for the introduction of key concepts and more general statements of that philosophy. In those texts, Marx emphasizes his commitment to materialism (as opposed to idealism), the role of contradiction and historical change (in the sense that capitalism and capitalist forms of thought are neither universal nor transhistorical), and the dialectical movement from the abstract to the concrete (as a way of producing a different, many-sided knowledge of the economic and social world). Scholars have discussed and debated the philosophy of Marxist economics, based on those texts and *Capital* itself, ever since.

Modernism

Traditional interpretations of Marxist economics are based on a modernist philosophy. As David F. Ruccio and Jack Amariglio discuss in some detail, modernist Marxist economists tend to invoke a "mirror of nature" epistemology and a deterministic notion of causality. For them, *Capital* presents a science of capitalism based on an objective labor theory of value (as against, e.g., the preference-based subjective theory of value of neoclassical economics). The problem of Marxist economics, on this interpretation, is to identify the order underlying the apparent disorder of capitalism and to point in the direction of a transition from capitalist disorder to the planned order of socialism.

Exchange: Prices/Profits
↑
Production: Value/Surplus value.

One illustrative example concerns the relationship between production and exchange. According to the modernist reading, the value and surplus value (the extra value created by laborers and appropriated by capitalists) created in the orderly process of capitalist production determine, and thus explain, the prices and profits that obtain in the chaotic and uncertain realm of capitalist exchange. Since the anarchy of a capitalist economy leads to crises, modernist Marxists argue that the disorder and irrationality of private property and markets can and should be replaced by the order and rationality of state property and planning within socialism.

Postmodernism

The postmodern interpretation is quite different. The theory of knowledge is relativist (Marx's approach, like all economic theories, is distinguished by a particular entry point, logic, and set of social consequences), and causality is nondeterministic (in the sense that everything is conceived to be both cause and effect, or "overdetermined"). The aim of Marxist economics, according to this philosophy, is to produce a particular class-analytical story about capitalism and to establish a nonexploitative

or communal way of organizing the economy and social life.

Production: Value/Surplus value → Exchange: Prices/Profits.

Again, the relationship between production and exchange can be used to illustrate the approach. On the postmodern interpretation, they exist on the same level. Thus, for example, capitalists need to purchase commodity inputs at their exchange value, which are then used to produce new commodities, which in turn need to be sold to realize the embodied value and surplus value. So each commodity has two numbers attached to it—production value and exchange value—each of which determines the other. Capitalism faces problems, related to class exploitation, that occur in both production and exchange (and the interaction between them). For postmodern Marxian economists, socialism is a way of eliminating capitalist exploitation (and its consequences) and creating a situation in which workers appropriate and distribute the surplus they create.

Critique

A third interpretation is best categorized as a *non-philosophy*. The idea is that *Capital* and other economic texts by Marx are distinguished less by a fully worked-out philosophy than by an approach emphasizing the critique of existing concepts and theories. Therefore, what is of interest today is the way Marx critically engaged dominant economic ideas. Since most of the categories that appear in Marx's economic writings (including the labor theory of value) are taken over from classical political economy, what is specifically Marxian (Marx's own) is the critique of the way the classical economists conceptualized those categories. Marx deconstructed the accepted understanding of those categories, taking what was natural or given by classical economics and denaturalizing them and making them social and historical. On this reading, *Capital* is not a Marxian theory of capitalism but a Marxian critique of mainstream economists' conception and celebration of capitalism—and thus a critique of capitalism itself.

This non-philosophy can also be illustrated in terms of the relationship between production and exchange. Marx began with the strongest assumptions of classical political economy—such as the ideas of fairness and equality—in order to show that the equality of production and exchange cannot but be accompanied by a fundamental unfairness: the extraction of labor from labor power. Laborers, under capitalism, get less in the form of wages than the value they create in production. Thus, Marx's idea of exploitation is both defined by and demonstrates the impossibility of capitalist conceptions of justice. Socialism, on this interpretation, represents a refusal of capitalist economic ideas—of production, exchange, justice, and so on—and a way of imagining and creating alternative (including collective or communal) forms of economy.

David F. Ruccio

See also Analytical Marxism; Capitalism; Causation in the Social Sciences; Determinism; Dialectic, in the Social Sciences; Heterodox Economics; Idealism; Marxism and Social/Historical Explanation; Marxist Ethics

Further Readings

Marx, K. (1904). *A contribution to the critique of political economy*. Chicago, IL: Charles H. Kerr.

Marx, K. (1973). *Grundrisse: Foundations of the critique of political economy* (M. Nicolaus, Trans.). New York, NY: Random House.

Marx, K. (1977). *Capital* (3 vols.; B. Fowkes, Trans.). New York, NY: Vintage.

Ruccio, D. F., & Amariglio, J. (2003). *Postmodern moments in modern economics*. Princeton, NJ: Princeton University Press.

Marxist Ethics

For more than a century, it has been well documented that Marx's relationship with the canonical philosophical terms of ethical inquiry is a troubled one. On one level of analysis, the troubled relationship runs parallel to the relationship between Marxism and philosophy, more generally. On another level, Marx and many Marxist thinkers have simply been skeptical of the role that ethical language plays in understanding and changing society.

This entry reviews the ambivalent nature of Marx's own pronouncement on ethics and of Marxist ethics as it developed, presents the main issues and

alternative interpretations, and raises a number of critical points.

Ambivalence

On the theoretical plane, the predominant line of Marxist thought has been to empty out philosophy into political economy. It is significant that Marx and much of Marxism followed the positivist paradigm in this. On the plane of practical struggle, Marx continually reminds us that "being determines consciousness," not the other way around; the lesson seems to be that people seek to change systems not first of all as a response to injustice but rather because they cannot live within the existing social parameters any longer. Under this paradigm, all struggle is reactive—and so ethical (and even political) concepts are epiphenomenal and insubstantial. Sometimes these concepts are even distinctly unhelpful, or they foster illusions; in Marx, there does not seem to be a difference between the ethical critique and "bourgeois moralizing," and all ethical language for Marx also carries with it theological overtones.

At the same time, Marx's writings are filled with ethical language; there are moments when he writes in the tonalities of an Old Testament prophet. However, this "ethical track" detected in Marx is clearly subordinate, and a great deal of Marxism since Marx has wrestled with the question of what to do with this subordinate track: Does one (1) excise it altogether, (2) recognize it but keep it in a subordinate position, (3) integrate it more fully with Marx's "scientific socialism," or (4) even raise it to a predominant position? Of Marxist thinkers since the time of Marx, a map of sorts can be made using these four possibilities. However, even understanding the nature of the "it" (the "ethical track," Marx's employment at times of ethical terms or terms with an ethical resonance, as, e.g., in his undisputed masterpiece, *Capital*, where Marx calls capitalists "vampires") is difficult.

Issues

One fruitful line of inquiry is to ask the following questions: Assuming that Marx does at times employ ethical language in a substantial and not merely ephiphenomenal or purely rhetorical way, is there a connection that can and should be made with thinking on ethics in the philosophical canon? Is it possible and fruitful to argue, for instance, that, insomuch as Marx's arguments have to do with ethics, this ethics most resembles the ethical theorizing of a particular figure or system?

Speaking very broadly, there are many things in Marx that find precedence in Judaism and Christianity. These systems of thought and forms of life, in turn, have been interpreted through the philosophies of Plato and Aristotle. Marx accepts the distinction in Plato between appearance and reality, though in terms of a materialist dialectic that has no need of a realm of forms. It could be argued that Marx accepts Aristotle's distinction, in the *Politics*, between caring for a household and the "mere getting of money." However, in Plato's and Aristotle's thinking, there is a central place for "the good" or "justice" as an "independent variable" as it were; such "metaphysical" and even "theological" tendencies are what Marx hopes to banish with a theory that is purely descriptive and not at all normative.

So, again, Marx's theory of labor, property, and wealth creation seems to have precedents in Thomas Aquinas and John Locke. However, there is a normative dimension in their shared sense of what is "proper" to a person—this has to do with making some part of the earth proper to oneself by mixing one's labor with the earth. In Aquinas, this notion again goes back to Aristotle, and it is the basis for the legitimate appropriation of that which, in nature, is the common inheritance of either all (even all creatures, not only humans) or, perhaps, none. In Locke, the further step is taken toward self-ownership (and the ownership of one's own body); a larger discussion of these issues would entail examining the shifts in culture and thought that allowed such a notion to gain traction. In Marx's early work, before *The German Ideology* (considered the first work of "scientific Marxism," written around 1845 but not published until 1932), Marx at least coexisted with (but sometimes actively embraced) the resonances of his work with strictly ethical terminology.

In his early work, injustice is manifest in alienation. While in his early work alienation is understood in a materialist framework, as the illegitimate expropriation of the fruits of an individual's labor (and, with the development of capitalism, the predominance of the strange "need for money," already an indication in Aristotle, Aquinas, and the Renaissance tradition of civic republicanism, of which Thomas Jefferson was one important inheritor, that something in society is increasingly out of

kilter), in the later work the notion of alienation is replaced by that of exploitation (specifically, exploitation of the ability to work, or "labor-power," as Marx called it), and this is given a purely mechanistic definition. Exploitation is a mechanism that, as it unfolds in the midst of other mechanisms (technical change, the tendency of the rate of profit to fall, the rising organic composition of capital, the increasing preponderance of "dead labor" over "living labor," the increasing impoverishment of the proletariat—but this last is not a category of ethical judgment), leads to crises of capital, and first of all a crisis of overproduction. None of this in Marx's view requires ethical categories—or, again, perhaps not even philosophical categories more generally, except perhaps for some concept of dialectics.

That exploitation—and the articulation of a society based on exploitation (of "labor-power")—is oppressive, that it requires political mechanisms that are those of domination, and that it requires ideological mechanisms that foster false consciousness (significantly, on the model of religion and with the help of "metaphysics") cannot be doubted, not even by Marx. Indeed, he opened up the examination of these mechanisms in a whole new way. But to discuss all of these things in ethical terms is beside the point. It is beside the point politically, economically, and socially, at best a distraction and mere sentimentality and at worst a fostering of dangerous illusions, another way of promoting *false consciousness*.

Coming at this from the other direction (which leads not to the breakdown of capitalism but rather to the formation of an alternative), the basis of working-class or proletarian solidarity *is not ethical*. This is a key point for Marx and for almost all that has called itself "Marxism": The basis of solidarity is not ethics but, rather, *interests*. As one might expect, granting a central place for the interests of the working class (which, of course, Marx does in a thoroughly systematic and rigorous way—these are not interests understood crudely or narrowly) brings Marx into proximity with utilitarianism.

Recent Alternative Interpretations

This line of analysis has led, in more recent years, to the interweaving of methodological individualism and rational choice theory with certain aspects of Marxism, as seen in the work of Jon Elster and others. This discussion was presaged in certain aspects by the debate in French Marxism in the 1960s on the question of humanism (Jean-Paul Sartre) and antihumanism (Luis Althusser). It is indicative to point out, in relation to our discussion of Marxist ethics, that Althusser and other structuralist Marxists advocated "theoretical antihumanism," an antiteleological, antihistoricist version of scientific Marxism, combined with "practical humanism." On the other hand, a different approach, the explicit articulation of Marxism as a form of utilitarianism, is found in the work of Steven Lukes and David Schweickart (the latter a leading proponent of market socialism).

There has been a great deal of discussion on the question of whether it is possible to get Kant and Marx on the same page, perhaps with the help of Aristotle. It is even conceivable to imagine anything recognizable of Marx or Marxism that is at the same time a form of moral realism or moral cognitivism in the broad sense of recognizing ethical questions as real and not simply epiphenomenal or merely ideological or rhetorical.

Contemporary Critical Challenges

The emergence, in the 20th century, of capitalism as an integrated, global system, one that takes up into itself the oppressive schemes of colonialism, patriarchy, racial hierarchy, and heteronormativity, makes these pressing questions. Two things seem clear by now. First, a calculus of interests alone will not close the gaps in such a way as to ground an effective and enduring solidarity. Second, for all that Marx was right to expose bourgeois moralizing (this is still an important aspect of ideology critique), even if there are moments when this exposure seems misdirected (as in the case of animal concerns, for example), it was wrong to go so far as to dismiss "the ethical" altogether.

One place to start with a correction of Marx on these points is a response to Hegel's dismissal of Kant's ethics as "empty formalism," a dismissal that Marx and Engels affirmed.

Finally, the term *Marxist ethics* also has another, somewhat different connotation: If one is a socialist, Marxist, or communist, what "ethics" should one have? Is there a "proletarian" ethics? Is there a sense in which ethics (and politics, for that matter) could be both "universal" and "partisan" at the same time? Marx urged that the proletariat must liberate itself and all humankind, too. Is there a way to contribute

to this emancipatory task that does not sacrifice ethics to "politics" or even to economics or revolutionary military tactics? Even though these questions have also been debated for more than a century, the challenge of the new century will almost certainly require new approaches. However, there remains much that deserves to be brought forward, even if recast in significant ways, in the tradition of ethical thought that Marx and Engels were too quick to leave behind.

Bill Martin

See also Alienation: From Philosophy to Sociology; Capitalism; Ideology; Marxism and Social/Historical Explanation; Marxist Economics; Moral Cognitivism; Neo-Marxism; Oppression

Further Readings

Cohen, G. A. (1995). *Self-ownership, freedom, and equality*. Cambridge, England: Cambridge University Press.
Elster, J. (1985). *Making sense of Marx*. Cambridge, England: Cambridge University Press.
Geras, N. (1997). *Discourses of extremity: Radical ethics and post-Marxist extravagances*. London, England: Verso.
Kain, P. J. (1991). *Marx and ethics*. Oxford, England: Oxford University Press.
Levine, A. (2003). *A future for Marxism? Althusser, the analytical turn, and the revival of socialist theory*. London, England: Pluto Press.
Lukes, S. (1987). *Marxism and morality*. Oxford, England: Oxford University Press.
MacIntyre, A. (2010). *Marxism and Christianity* (3rd ed.). London, England: Duckworth.
Martin, B. (2008). *Ethical Marxism: The categorical imperative of liberation*. Chicago, IL: Open Court.
Schweickart, D. (2011). *After capitalism* (2nd ed.). Lanham, MD: Rowman & Littlefield.
Wolff, R. P. (1988). *Moneybags must be so lucky: On the literary structure of Capital*. Amherst: University of Massachusetts Press.

Mathematical Models, Use in the Social Sciences

Mathematics, in general, is widely used in social science disciplines such as economics, sociology, and political science. Mathematical models, in particular, are used to study phenomena that are very different in nature, such as individual economic behavior, population growth, and aggregation in urban areas. The branches of mathematics used in these models are also very different: They span differential equations, matrix calculus, computational algorithms, probability, and statistics.

This entry focuses on the use of probability and statistics in "structural models." These are models used in quantitative (causal) analysis in, for example, econometrics or demography. One reason to focus on structural models is the vivid debate they generate, especially with respect to the notions of *causation*, *intervention*, and *mechanism*—notions central to the social sciences.

Probability theory and statistics offer tools to analyze random phenomena, such as coin tossing or waiting time in a queue, or more complex social phenomena, such as migration behavior or changes in mortality, morbidity, and fertility of populations. Probability and statistics are very useful in the analysis of data collected through censuses, surveys, or other types of methods. Data are observations about a number of different aspects of populations or individuals and are organized into *variables*. There are different types of variables that can be categorized according to the following criteria:

Genre and scale: continuous/discrete, quantitative/qualitative

Role: explanatory/response, observed/latent/instrumental/proxy

Level: individual/aggregate

Field: socioeconomic/demographic/biological/and so on

Quantitative analysis has a long tradition, starting with the pioneering and seminal works of the demographer and astronomer Adolphe Quetelet and the sociologist Émile Durkheim. Major improvements have been done by, to name a few, Sewall Wright and Otis Dudley Duncan, up to the most recent advancement by econometricians such as Jim Heckman or Kevin Hoover and computer scientists such as Judea Pearl or Peter Spirtes, Clark Glymour, and Richard Scheines. What is peculiar to this development is that while the early methodologists overtly and explicitly adopted a causalist perspective, the most recent generations have shown some skepticism as to whether we can infer causation from statistics.

The use of probability and statistics in the social sciences has consequences for the kind of explanation that the models offer. Statistical explanations

are partial and incomplete in the sense that no phenomenon will be "fully" explained. Any explanations will carry a certain amount of indeterminacy. However, this is not necessarily a statement about the social phenomena themselves but just about our *knowledge of them*.

Quantitative analysis can be used to provide a description of a social phenomenon (*associational models*) or to go beyond description and to find out its causes (*causal models*). More precisely, associational models study how one variable changes in response to changes in other variables. For instance, a number of studies in the 1960s revealed that in developing countries, variations in child survival rates were constantly associated to variations in levels of maternal education: Child survival was higher for higher levels of maternal education. Causal models try to go beyond associations by imposing further constraints that allow interpreting associations causally. For instance, to follow up on the same example, in the 1980s, a more structured model to explain the association between mother's education and child survival was proposed. This model introduced a structure of proximate biological factors (e.g., nutrient deficiency) and more remote social factors (access to infrastructures) that explained the intermediate steps between mother's education and child survival. In addition to adding "theoretical" support to the associations, causal models have a more sophisticated statistical apparatus.

Both associational and causal models share a minimum set of "technical features." Notably, both kinds of model make assumptions about some characteristics of the variables (typically, normality, linearity, nonmeasurement error, and noncorrelation of error terms.). The apparatus of causal models is, however, richer, and this is what allows us to interpret models causally. In particular, causal models (ought to) (a) make background knowledge explicit, (b) specify the conceptual hypothesis to put forward for empirical testing, and (c) make a number of assumptions that go beyond statistics, for instance, about causal ordering of the variables or the underlying mechanism, about the structure of the causal relation, or about sufficiency (viz., that all and only the relevant factors have been included in the model).

Another important difference between associational and causal models lies in the tests they perform on the joint variations between variables of interest. Associational models are mainly concerned with ensuring that co-variation is not spurious. Causal models go much further, instead, and perform a number of additional tests. Tests for invariance, for instance, are meant to ensure that the (putative) causal relation is sufficiently stable across different partitions of the data set or under suitable interventions. Tests for exogeneity are instead meant to ensure that causes and effects are correctly "separated" and therefore the causal structure is correct. This last test is performed against the "recursive decomposition," which breaks down an initial joint probability distribution into "smaller" pieces.

Two issues are worth emphasizing. First, according to some interpretations of causal models, the recursive decomposition may be viewed as representing a mechanism underlying the data (also called *data-generating process*). Thus, the recursive decomposition carries explanatory power insofar as it gives the "functioning" of the underlying mechanism. Second, the condition of invariance is a matter of controversy. According to one school of thought, invariance has to be tested against manipulations (whether actual or ideal) of the putative cause. But according to another view, invariance doesn't need to be tested against manipulations. It is often the case in the social sciences that manipulations are unfeasible or meaningless. In such cases, invariance is simply tested across chosen partitions of the population.

There is an important caveat. The mathematization of social research certainly is a positive step forward. Mathematization allows great precision in analysis, and it allows comparisons between different studies. However, it is questionable whether this mathematization, *alone*, solves all the problems related to causal inference, explanation, or prediction. Making background knowledge explicit looms large in making social studies as objective as possible. This also means having an overarching view of modeling, where statistical testing is but one part.

Federica Russo

See also Causal Explanation, in Philosophy of Science; Causation in the Social Sciences; Econometrics: Methodological Issues; Models in Science; Models in Social Science; Probability

Further Readings

Franck, R. (Ed.). (2002). *The explanatory power of models*. Dordrecht, Netherlands: Kluwer.

Freedman, D. A. (2005). *Statistical models: Theory and practice*. Cambridge, England: Cambridge University Press.
Kemeny, J., & Laurie, S. J. (1978). *Mathematical models in the social sciences*. Cambridge: MIT Press.
Russo, F. (2009). *Causality and causal modelling in the social sciences: Measuring variations*. New York, NY: Springer.

Mechanism and Mechanismic Explanation

This entry introduces the notion of mechanism in relation to that of system and shows the relevance of mechanism to explanation in both the natural and the social sciences.

Most things in nature and in society are composite: Only the elementary particles in empty space are simple, and even so just for a while. There are two kinds of composite things: agglomerates, such as sand dunes and spontaneous crowds, and systems, such as molecules and families. Unlike agglomerates, systems are structured.

System and Mechanism

A *system* can be identified by its composition or set of parts, structure (or set of relations among its parts), and environment or context. This description suffices for systems of symbols or of concepts, such as sonnets or proofs, but not for concrete systems, such as crystals or governments.

If concrete, a system has also one or more mechanisms, or characteristic processes. For example, metabolism and cell division are mechanisms of most living cells, and trade and competition are market mechanisms. In sum, a concrete system can be modeled by the ordered quadruple *composition–environment–structure–mechanism*.

The first two components of this quadruple suffice to characterize an agglomerate and the first three, a symbolic system. The five concepts in question should enjoy pride of place in any metaphysics, but they don't, because most metaphysicians care more for possible worlds than for the real world. In fact, the term *mechanism* has been practically absent from philosophy.

True, mechanisms have made a sort of cameo appearance in the recent philosophy of science literature, but in a loose way and, in particular, detached from system and, thus, remote from the usage of "mechanism" in science and technology. In these fields, it has been ubiquitous since the Scientific Revolution. Let us look at a few examples.

The concepts of system and mechanism are likely to have been used tacitly by craftsmen and engineers since remote antiquity. But they emerged explicitly only around 1600 in both astronomy and physiology: Recall the expressions "planetary system" and "cardiovascular system." Before that time, the Sun and its planets had been treated in isolation from one another, and only Newton provided the cement—gravity—that held those bodies together into a system. Likewise, before William Harvey, the organs of the human body were viewed separately from one another, and nobody knew that the heart's contractions and swellings kept the blood circulating.

Note that neither gravity nor the heart's systoles and diastoles are mechanical mechanisms. In fact, throughout modernity, the original concept of a mechanism as a mechanical contraption was silently broadened to all the fields of knowledge—unbeknown to most philosophers.

Here are a few examples of nonmechanical mechanisms. (a) The impressed electric field (voltage) moves the electric charges around a circuit, which in turn generate a magnetic field that opposes the current (self-induction). (b) Two of the more prevalent mechanisms that explain chemical reactions are electron transfer (electrovalent bond) and electron sharing (covalent bond). (c) Darwin explained evolution by two main processes: inheritance with modification and natural selection—both of which proceed from birth. (d) Aversive memories are extinguished not by the superego but by the cannabinoids secreted by our own bodies: These molecules wreck the neuronal process in the organ that stores such memories—the amygdala. (e) The central banks implement metamechanisms, such as currency control, to control the trade mechanisms. (f) Dogmatic education, bribing, censorship, and repression are mechanisms for checking the invention of ideas and practices feared to destabilize the mechanisms designed to prop up the powers that be.

Because a number of mechanisms may operate in parallel in one and the same system and because some of them may interfere with one another, it is convenient to distinguish essential from nonessential mechanisms. The former are those peculiar to systems of a certain kind, whereas the latter may also occur in systems of different kinds. For example,

contraction is essential to a muscle but inessential to a cell, and loaning money is essential to a bank but optional for a manufacturer.

We are now ready to propose and refine this definition: An *essential mechanism* of a system is its peculiar functioning or activity. In other words, an essential mechanism is the specific function of a system—that is, the process that only it and its kind can undergo. So much for mechanisms; let us now explore their role in explanation.

Explanation and Mechanism

The disclosure of mechanism is central to explanation. In fact, this is what counts as explanation in the sciences and technologies: to find out how things work and what their mechanisms or modi operandi are. In contrast, the pseudoscientists do not look for mechanisms. For example, parapsychologists do not explain how psychokinesis (moving things by sheer mental power) works, and believers in homeopathy do not exhibit the mechanism whereby a highly diluted solution, such as one with a single molecule of the "active principle" per galaxy, could cause or inhibit a biological process.

A similar reasoning is used to evaluate inventions: No patent is ever granted unless the inventor succeeds in explaining how the novel device works. This is why the most effective way for a patent office to deny a patent on an allegedly revolutionary design is to point out that the proposed mechanism is incompatible with well-known laws, such as the law of conservation of energy. No law, no possible mechanism; and no mechanism, no explanation. No wonder then that the hallmark of modern science is the search for the mechanisms behind facts, rather than the mindless search for data and the statistical correlations among them.

The standard conception of explanation in terms of law(s) and circumstance may be called *subsumptive*, as it subsumes the particular under the general. A *mechanismic* explanation, by contrast, resorts to law statements referring to mechanisms.

Here is an example from sociology and management science: "The inertia [resistance to change] of a social system is proportional to its size." This explains why even friendly takeovers, which require quick adaptations, are hazardous to corporations. In turn, the relevance of size to inertia is explained by the need for face-to-face (or at least screen-to-screen) contacts to maintain the cohesion of the system and, thus, ensure its behaving as a unit.

To put it schematically, we have split the initial statement "↑Bulkiness ⇒ ↑Inertia" into "↑Bulkiness ⇒ ↓Contacts" and "↓Contacts ⇒ ↑Inertia."

This argument is clarified when expressed with the help of the standard symbolism of elementary logic. We started with a law statement of the form "$\forall x(Ax \Rightarrow Bx)$" and analyzed it as the conjunction of hypotheses of the forms "$\forall x(Ax \Rightarrow Mx)$" and "$\forall x(Mx \Rightarrow Bx)$," where M refers to a key feature of some mechanism. When handling a particular fact, such as the failure of a megacorporation that resulted from the merger of two successful businesses, we must conjoin the former premises with information about the increase in size.

To summarize, look for system and explain by reference to mechanism.

Mario Bunge

See also Explanation, Theories of; Feedback Mechanisms and Self-Regulatory Processes in the Social Sciences; Laws of Nature; Pseudoscience; Systems Theory

Further Readings

Bunge, M. (1967). *Scientific research* (Vol. 2, chap. 9). New York, NY: Springer.

Bunge, M. (2004). How does it work? The search for explanatory mechanisms. *Philosophy of the Social Sciences, 34,* 182–210.

Machamer, P., Dardren, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67,* 1–25.

Merton, R. K. (1968). *Social theory and social structure* (3rd ed.). New York, NY: Free Press.

Mereology: Parts and Wholes

The relationship of part to whole is familiar and universal. Like the relationship between a house and its front door, an elephant and its trunk, and Beethoven's *Eroica* and its first chord, we find part–whole relations all around and grasp the notion intuitively from infancy. Despite the ubiquitous applicability of the concept of part, and its not infrequent intrusion into philosophical discussion, it was not until the 20th century that philosophers and logicians progressed from an intuitive understanding to the first formal

axiomatic theories of part and whole. The general idea was envisaged but not executed by Edmund Husserl in 1901; it was then executed sketchily by Alfred North Whitehead and precisely by Stanisław Lesniewski in 1915. The axiomatic theory of part and whole became generally known to the English-speaking logico-philosophical world in 1940 under the title "calculus of individuals," after a paper of that name by Henry S. Leonard and Nelson Goodman. From the 1980s onward, it entered the toolkit of postpositivist ontology and is now an indispensable component of modern analytic metaphysics and its applications. The theory of part and whole is generally known by the name *mereology*, given to it in the 1920s by Lesniewski, from the Greek *meros*, "part."

This entry elucidates the core formal characteristics of the part relation, indicates what formal properties have been ascribed to it beyond this core and why, and considers what variations and extensions have been proposed concerning the part relation. We will look at several philosophical controversies that revolve around part–whole relations, before considering crucial applications of mereology outside philosophy. Mereological themes can be found in many different philosophical subfields, but the philosophical foundation of part–whole relations is of particular importance to the philosophy of the social sciences in view of the holism/individualism debate.

Core Formal Properties

The basic notion of one thing being a part of another is so familiar that it hardly needs introduction. But on reflection, we find that for a relation to be a part relation, it has to fulfill a small number of formal (logical) requirements. These may be set out as follows:

Existence: Nothing can be a part of another thing unless both things exist.

Irreflexivity: If one thing is a part of another, they cannot be the same thing.

Transitivity: If one thing is a part of another and the second thing is a part of a third, then the first is also a part of the third.

Supplementation: If one thing is a part of another, then the larger thing must have an additional part that is not shared with the smaller.

These requirements may be set out with a modicum more formality using variables and logical notions as follows, where the modal component of the above vernacular statements is transferred to the status of the principles as statements.

Existence: For all a and b, if a is a part of b, then a exists and b exists.

Irreflexivity: For all a and b, if a is part of b, then b is not identical with a.

Transitivity: For all a, b, and c, if a is a part of b and b is a part of c, then a is a part of c.

To prepare for the next requirement of supplementation, we define a couple of auxiliary notions:

Ingredient: a is an ingredient of b iff (definition) a is a part of b or a is identical with b.

Overlap: a overlaps b iff (definition) for some x, x is an ingredient of both a and b.

Supplementation: For all a and b, if a is a part of b, then for some x, x is a part of b and x does not overlap a.

The formal principles can be, and usually are, clothed in symbols, but they make no difference to the meaning intended.

Any relation that purports to be a part relation has to satisfy these principles, which can be taken to be analytic of the part relation. Other relations may by chance fulfill the principles, but only a relation that satisfies them by necessity is a part relation. Other than elucidating these formal properties and giving examples, the notion of part cannot be further explicated. While the notion of part is ontological rather than logical, it comes as close to being logical as an ontological relation can be. It is explicated by its formal properties. It is thus one of the principal notions of formal ontology.

Beyond the Core

So far we have expounded only the formal core of the part relation. But in many or most formal treatments, additional properties are ascribed to it. The two most important and frequently invoked properties are as follows:

Extensionality: Any things having parts and having the same parts are identical. For all a and b, if a and b both have parts and every part of a is a part of b and vice versa, then a is identical with b.

Summation: For any collection of (existing) objects *A*, there exists an object consisting of the objects in *A* and nothing else. For all *A*, if all of *A* exist, then there exists an object *s* such that every one of *A* is a part of *s* and every part of *s* overlaps one of *A*.

Of these, the former, the principle of (mereological) extensionality, is by a wide margin the less controversial. Most formal theories of part and whole assume it. In many contexts, it is self-evidently correct, so that finding potential counterexamples is not easy. The most likely counterexamples concern cases involving a material object and the matter of which it is made. For example, a clay statue and the lump of clay of which it is made seem to have the same parts, yet they may not be the same thing, because the clay may survive distortion and separation but the statue may not. However, there are theories that proclaim the clay and the statue to be identical and other theories that proclaim them to be different but not having the same parts. For example, if the clay predates the statue, then it is said that it has temporal parts that the statue lacks. Both of these strategies enable the Principle of Mereological Extensionality to be preserved: Both are metaphysically controversial. This offers an initial indication that mereological theory may go to the heart of metaphysics.

The Summation Principle is relatively much more controversial. It states that any existing objects whatever compose an individual, and for this reason, it is sometimes called the Principle of Universal Composition. Where objects are widely separated in space and time and/or come from diverse ontological categories, it is relatively less plausible than when objects form a relatively homogeneous category, as, for example, under forms of materialistic nominalism. We shall return to consider Summation below.

The motivation for the additional principles of Extensionality and Summation derives from the context in which modern formal mereology was born. In Leśniewski, mereology served as a nominalistic ersatz for set theory, and for this reason, it was expedient that its principles be as strong as possible, in fact amounting to those of an extended Boolean algebra minus the null element. In the case of Whitehead, mereology was formulated to serve as an acceptable ontological basis for physical geometry. In its first embodiment, Whitehead's mereology applied to a domain of four-dimensional events, and in its second, it applied to the spatiotemporal regions occupied by events. Whitehead did not accept the Summation Principle, since one of its consequences is that there is a unique maximal individual, the sum of everything, whereas he considered that everything both is and has a part (for the opposite view of *mereological nihilism,* see the entry on Nihilism). Nevertheless, the homogeneity of his domains means that Extensionality is unproblematic for him, as for Leśniewski. Away from such mathematical motivations, and in particular in application to everyday objects, the noncore principles lose a good deal of their plausibility. Nevertheless, there have been many metaphysicians who accept the strongest theory, that of Leśniewski, which is usually known as Classical Extensional Mereology. Much ingenuity has gone into the investigation of different primitives and axiom systems, all resulting in what is from a metaphysical point of view the same theory.

A further consequence of Whitehead's principle that everything has a part is anti-atomism. A mereological atom is an object that has no parts. Therefore, Whitehead considers that there are no atoms. By contrast, atomism states that everything is completely composed of atoms. Leśniewski's mereology leaves the question of atomism open: It can be extended by atomistic, anti-atomism, or indeed a third hybrid position. Empirical considerations also seem incapable at present of resolving the question of whether the world is atomistic or not.

Metaphysical Controversies

Many metaphysical disputes have a mereological dimension. For example, there is the question of whether substances can have parts (Leibniz: No) or whether substances can have other substances as parts (Spinoza: No, because only the biggest whole is a substance; Aristotle: No, because the parts of substances are not substances; Descartes: strictly No, because God is the sole substance and has no parts, but nonstrictly Yes, because spatially extended bodies are substances and can have other such bodies as parts).

Some metaphysical controversies, however, turn directly on mereology, most especially the question as to when a collection of things mereologically compose another thing. This is known, following Van Inwagen, as the Special Composition Question. Here, several kinds of answer can be and have been given: Never, because all that exists are atoms and they never compose anything (mereological nihilism); Never, because there is only the whole and it has no parts (Parmenidean monism); Always

(mereological universalism); and, finally, Sometimes but not Always. The last kind of answer breaks down into types according to the condition under which smaller things compose larger things. Van Inwagen himself proposes a fairly extreme view, *organicism*, according to which, things only compose a larger thing when the smaller things are atoms and the larger thing is an organism. So table legs as parts of tables and hands as parts of humans are out, in the first case, because neither table legs nor tables exist, and, in the second, because hands do not exist though humans do. According to common sense, things sometimes, but not always, compose larger things, but common sense harbors no principles as to when or why.

One reason offered in support of universalism is that it allows a simple answer to the question as to the relation between a composite whole and the complete collection of its parts or, alternatively, of its atomic parts. The thesis of composition as identity says that the whole simply is identical to this collection: They are it, and it is them. The disadvantage of this position is that it makes one thing, the whole, identical to many things, the parts taken together, which some consider to violate the logic of number.

An older mereological controversy with roots in medieval discussion is whether an object has its parts essentially. An object *a* has a part *b* essentially if *a* could not exist and fail to have *b* as a part. The view that every object has all its parts essentially is known as mereological essentialism. The most prominent modern representative of this view is Roderick Chisholm. The view has to explain the common-sense observation that most physical objects, such as stars, organisms, and artifacts, may change their parts over time and might have had other parts than those they do have. The usual way around this for mereological essentialists is to claim that only metaphysically second-rate objects flout essentialism, while "genuine" objects satisfy it. Away from the controversy, it remains of interest whether any objects satisfy essentialism. Events and regions of space and time seem good candidates, while physical objects do not. For the latter, one may still enquire which if any of an object's parts are essential to it: So to an atom, its nucleons appear essential; to a man, his hair is inessential; while for cases like a human's brain, it is not clear. A virtue of essentialism is that it answers the truthmaker problem for true part–whole propositions, the question as to what object or objects simply by existing necessitate the truth of the proposition. In a case of essentialism, the simple answer is "the whole." For nonessentialist cases, another and as yet untried solution must be found.

Variants and Extensions

Mereology is typically pursued logically as a theory of one binary relation, part–whole, and its cognates. But in application, additional factors enter in, notably time. Where an object like a car undergoes repair involving the replacement of one part by another, the simple binary relation generally does not specify the additional factor, which concerns when certain objects are parts of others. A part may, for example, be transitory or temporary, or it may be permanent, part of its whole as long as the whole exists. It may be an initial part, a terminal part, or a part the whole has partway through its life. To formulate such propositions, one must either add tense to the mereological propositions or relativize the part–whole relation to time. By adding the temporal specification "at *t*" to each of the predicates of the Core Mereology, or also its extensions, one obtains mereology for a time, or Synchronic Mereology. For example, a piston cannot be a part of a car at a time unless both piston and car exist at that time. Similar is the case for the other principles. Considerations of time, however, tend to undermine compositional universalism. It stretches credulity to envisage a whole consisting solely of Napoleon's left foot and Bismarck's moustache, which ceases to exist at the death of Napoleon and comes back into existence with the first labial hirsute growth of the Iron Chancellor.

Synchronic Mereology offers no guidance on what to do about identity through time, such as whether a certain object is required to have certain others as permanent parts or how the gain and loss of parts might affect the existence and identity of wholes. One notorious metaphysical conundrum concerns the fabled *Ship of Theseus*, which over time had all its parts replaced, so that philosophers disputed whether after the change it was the same ship or not. Thomas Hobbes compounded the agony by supposing the original parts to be saved and then reassembled, giving rise to two competing candidates to be considered as Theseus's ship. Mereology is needed to formulate the problem, but it does not on its own provide a solution. For mereological

essentialists, for example, the ship ceases to exist the moment the smallest part is removed, but this answer is too remote from standard practice and language to command assent, except among certain metaphysical purists.

Mereological essentialism requires modal notions for its formulation, whether necessity and possibility applied to whole sentences or essentiality applied to predicates. However, beyond the formulation of concepts of essential and accidental parts and wholes and the discussion of the merits of mereological essentialism modal mereology has remained rather undeveloped.

Collection and Mass

Mereology standardly concerns itself with relations among individuals: the leaf and the tree, for example. But formally, exactly analogous constructions apply to collections consisting of more than one individual. The left-handed children in a school class form part of the class, and in many contexts, "part of" and "some of" are interchangeable. A single child is like a mereological atom: No part of it is one of the collection of children in the class. Questions of essential and accidental parts apply to collections in the same way as they do among individuals. This use of mereology is clearly of interest to sociobiology and social science, where groups of people or other animals are under consideration. Likewise, the ontology of part and whole applies to masses, in particular to masses of matter: The gin in a gimlet is part of the cocktail, though it is not an individual but a mass of stuff.

Mereological Vagueness

Mereology has traditionally been pursued in logic on the assumption that of any two objects *a* and *b*, *a* is either a part of *b* or not and there is no third case. But in real life, we often come across cases where it is unclear, for whatever reason, whether *a* is a part of *b* or not. This applies especially to changing things, such as organisms, stars, rivers, and social groupings. The thought that it may not be determinate what *a*'s parts are gives rise to the idea of ontic vagueness, vagueness of the identity of objects. The nature of vagueness is a disputed territory in philosophy, but whichever solutions are offered, they need to apply plausibly to purported cases of ontic vagueness.

Kinds of Part

When discussing ontological categories, such as processes and material objects, or collections and masses, it is fair to assume homogeneity of parts and wholes: Parts of material objects are material objects, parts of processes are processes, and so on. A second perspective is afforded bydiffering kinds of part. We have already encountered essential and accidental, temporary and permanent, atomic and composite, but most of the parts in which we take natural cognitive and scientific interest come in different kinds. This is particularly important in applications and will be discussed below. The fact that we take special interest in certain kinds of part, functional parts in organisms and artifacts, for instance, allows mereology to serve as a general ontological framework within which questions of the nature and working of composite items from many categories can be debated. These varieties of kinds of part can be traced without compromising the univocity of the part relation among individuals, and they afford some reason for compositional universalists to claim that while their arbitrary sums may not be interesting, they are at least consistent and harmless and offer an encompassing framework for all the kinds of part in which we are interested.

Applications

Its universality renders mereology supremely applicable, and it automatically enters into anatomy, physical geography, engineering, and social theory, to name just some obvious areas. In anatomy, the delineation and description of parts and their layout is what the science is about, while physical geography is the anatomy of geophysical parts as well as less clearly bulky features, such as valleys and cliff faces. In engineering, the whole point is to construct artifacts from parts that can be suitably manufactured and assembled and that when in situ and in action perform their intended function. And sociology is centrally concerned with different parts of society and smaller social groups, such as classes, age-groups, institutions, migrants, children, and so on. In all of these cases, mereology offers its neutral conceptual framework, and as the electronic acquisition, storage, and manipulation of increasingly complex data from such areas increase, the use and proper formulation of mereological concepts and principles in the relevant databases acquire increased

importance. In future, mereology will be of greater interest to practitioners than to logicians.

Peter Simons

See also Essentialism; Holism, in the Social Sciences; Kinds: Natural Kinds Versus Human; Supervenience Systems Theory

Further Readings

Burkhardt, H., Seibt, J., & Imaguire, G. (Eds.). (in press). *Handbook of mereology*. Munich, Germany: Philosophia-Verlag.

Chisholm, R. (1973). Parts as essential to their wholes. *Review of Metaphysics, 26,* 581–603.

Chisholm, R. (1975). Mereological essentialism: Further considerations. *Review of Metaphysics, 28,* 477–484.

Fine, K. (1999). Things and their parts. *Midwest Studies in Philosophy, 23*(1), 61–74.

Fine, K. (2010). Towards a theory of part. *Journal of Philosophy, 107*(11), 559–589.

Leonard, H. S., & Goodman, N. (1940). The calculus of individuals and its uses. *Journal of Symbolic Logic, 5,* 45–55.

Leśniewski, S. (1992). On the foundations of mathematics. In *Collected Works* (pp. 174–382). Dordrecht, Netherlands: Kluwer.

Simons, P. M. (1987). *Parts: A study in ontology*. Oxford, England: Oxford University Press.

Simons, P. M. (2006). Real wholes, real parts: Mereology without algebra. *Journal of Philosophy, 103*(12), 597–613.

Smith, B. (Ed.). (1982). *Parts and moments: Studies in logic and formal ontology*. Munich, Germany: Philosophia-Verlag.

Varzi, A. (2008). The extensionality of parthood and composition. *Philosophical Quarterly, 58*(230), 108–133.

Whitehead, A. N. (1919). *An enquiry concerning the principles of natural knowledge*. Cambridge, England: Cambridge University Press.

Metacognition and Agency

Metacognition of agency is defined as the person's knowledge of the extent to which the self is in control. The ability to know about one's own causal impact on the world and on other people and to make attributions of self-agency accurately is essential in many domains, including the attribution of credit and blame. As such, it has legal implications. The cognitive characteristic that most poignantly characterizes the insanity seen in schizophrenia is the individual's loss of this ability to know whether or not the self was the agent—whether he or she did something or whether it occurred as a result of an external agent or accidental force.

The ability to exercise control over one's actions—to both be an intentional agent and to know that one is—is at the heart of what is meant to be an autonomous, self-controlling human being. Metacognition of agency is related to people's perceptions of intentionally willing or planning an action and then causing that action to happen. As such, it is inevitably linked to questions of free will and intentionality. Some researchers have argued that because people can be radically wrong in their judgments of agency—believing that they have done things when they have not and have not done things when they have—the construct of free will is suspect. Despite the importance of agency, the ontological status of this philosophical and psychological construct is still open to debate. Most scientists agree that the intention, at least for purposes of scientific investigation, cannot be considered as something separate from the brain or as a nonphysical entity that, nevertheless, causes a physical brain/body reaction. The mind does not tell our brain to do something that is then actualized materially. But, at the same time, there is unequivocal support for the idea that brain-based top-down processing, whereby actions are instigated by and modulated in conjunction with plans and expectations, is a fact of human mental life.

Cognitive scientists sidestep the issue of whether, when apparent intention–action correspondences do occur, they do so because people have free will, and they instead focus on the factors that contribute to people's feelings of being in control. The uncontested conclusion that has emerged is that knowledge of one's own agency—rather than being direct knowledge that is inherently and infallibly available to the individual—depends instead on particular cues that people use to make their judgments. Some of these cues are inherently diagnostic, while some are not.

The most diagnostic, and extensively studied, of these cues is the potential discrepancy between the individual's plans and expectations, and the outcome. According to this scheme, when a person has a goal, it gives rise to an internal model of his or her intentions (called the inverse model) and expectations (called the forward model) about achieving the goal. This initiates a motor plan that provides the specifications about what needs to be done to achieve the

goal. The plan runs simultaneously with the person's motor actions. A comparator mechanism evaluates the correspondence of the actions and the plan. When the two streams match, the person is in control; when they mismatch, it is a usable cue that the person is not in control. A mismatch can also trigger a motor correction. This mismatch signal could be processed unconsciously. Indeed, some studies have provided evidence that it sometimes is unconscious insofar as appropriate motor adjustments are made without conscious awareness about the nature of the adjustment. In some circumstances, though, it can provide a diagnostic signal to a metacognitive monitor concerning whether or not the person was in control. Conditions, such as when external interference or turbulence is systematically manipulated to distort a movement, have been used to allow evaluation of the use of a discrepancy cue in the judgment process. When such a mismatch signal occurs, people reliably report that they feel out of control.

This diagnostic "mismatch" cue is not the only cue that contributes to people's judgments of agency, however. Judgments of agency are also affected by a delay in the effect of people's actions. Furthermore, nondiagnostic cues also contribute to these judgments. Judgments of agency are strongly influenced by people's perception of performance. When outcomes are good, people tend to say that they were in control; when outcomes are bad, they tend to say that they were not in control, regardless of the extent of their actual control. Patients with schizophrenia use this nondiagnostic cue in making agency judgments, to the exclusion of the diagnostic "mismatch" cue. The mirror neuron system, too, appears to contribute a spurious cue by providing internal evidence that other people's actions are one's own. The pervasively seen claim that individuals are disproportionately responsible for shared tasks suggests that people are inclined to take credit for other people's actions as well as their own. Finally, judgments of agency are influenced by feelings of reward. For example, Matthew Kirkpatrick and colleagues found that when methamphetamine users received the drug as compared with a placebo, their feelings of being in control were inflated, even though their performance was not. Young children also respond to artificially induced increases in performance as if they were increases in their own agency, whereas young and older adults are less likely to do so.

While these cues may contribute to people's judgments of agency, the self-referential judgment is itself distinct from the cues that provide its informational basis. The cues alone, as noted above, could exert control over behavior without participants becoming aware of them or their implications for the role of the self in action. Indeed, it is highly likely that nonhuman primates have in place many of the systems that provide the cues that people use to assess their own agency. They control their complex movements with remarkable grace, they respond to reward, they can evaluate performance, and they have a mirror neuron system. But while the underlying cues are almost certainly present, what is unknown is whether nonhuman primates also have a sense of self and the capability to relate these cues in a self-referential way to produce full-blown metacognition of agency. In humans, though, although their use can sometimes be faulty, these underlying cues are used to make self-referential judgments about the individual's own personal responsibility for causing an action.

Studies that have investigated human brain correlates of metacognition of agency underline the distinction between the ongoing cues to agency that occur while the individual is performing a task and the self referential judgment that occurs retrospectively. Many studies have shown the involvement of an area in the right parietal junction in particular, which is related to processing of a discrepancy introduced into a motor task. Studies have also shown evidence of caudate-reward system involvement when movements are running off smoothly according to plan and people are "in control." However, the activation during the judgment phase is in a different location from those seen during task performance—an area of the frontal cortex (BA10) that has been shown to be related both to metacognitive and to self-referential processing.

Janet Metcalfe

See also Agency; Free Will, Philosophical Conceptions of; Intention, Social Psychology of; Introspection (Philosophical Psychology); Mirror Neurons and Motor Cognition in Action Explanation; Neural Hermeneutics; Self-Knowledge; Social Cognition; Social Neuroscience; Unconscious Social Behavior

Further Readings

Carruthers, P. (2011). *The opacity of mind: An integrative theory of self-knowledge*. Oxford, England: Oxford University Press.

Decety, J., & Lamm, C. (2007). The role of the right temporo-parietal junction in social interaction: How low-level computational processes contribute to meta-cognition. *Neuroscientist, 13,* 580–593.

Dewey, J. A., Seiffert, A. E., & Carr, T. H. (2010). Taking credit for success: The phenomenology of control in a goal-directed task. *Consciousness and Cognition, 19,* 48–62.

Fleming, S. M., Weill, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science, 329,* 1541–1543.

Frith, C. D., Blakemore, S., & Wolpert, D. M. (2000). Interactive report explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews, 31,* 357–363.

Haggard, P. (2008). Human volition: Towards a neuroscience of will. *Nature Reviews Neuroscience, 9,* 934–946. doi:10.1038/nrn2497

Jeannerod, M. (2009). The sense of agency and its disturbances in schizophrenia: A reappraisal. *Experimental Brain Research, 192,* 527–532. doi:10.1007/s00221–008–1533–3

Jenkins, A. C., & Mitchell, J. P. (2010). Medial prefrontal cortex subserves diverse forms of self-reflection. *Social Neuroscience, 6,* 211–218. doi:10.1080/17470919.2010.507948

Kirkpatrick, M., Metcalfe, J., Greene, M., & Hart, C. (2008). Effects of intranasal methamphetamine on metacognition of agency. *Psychopharmacology, 197,* 137–144.

Metcalfe, J. (in press). Knowing' that the self is the agent. In H. Terrace & J. Metcalfe (Eds.), *Agency and joint attention*. New York, NY: Oxford University Press.

Metcalfe, J., Eich, T. S., & Castel, A. (2010). Metacognition of agency across the lifespan. *Cognition, 116,* 267–282. doi:10.1016/j.cognition.2010.05.009

Metcalfe, J., & Greene, M. J. (2007). Metacognition of agency. *Journal of Experimental Psychology: General, 136,* 184–199. doi:10.1037/0096–3445.136.2.184

Metcalfe, J., & Son, L. K. (2012). Anoetic, noetic, and autonoetic metacognition. In M. Beran, J. L. Brandl, J. Perner, & J. Proust (Eds.), *The foundations of metacognition* (pp. 289–301). Oxford, England: Oxford University Press.

Metcalfe, J., Van Snellenberg, J. X., DeRossa, P., Balsam, P., & Malhotra, A. K. (in press). Judgments of agency in schizophrenia: An impairment in autonoetic metacognition. *Proceedings of the Royal Society of London: B, 367,* 1391–1400.

Miele, D. M., Wager, T. D., Mitchell, J. P., & Metcalfe, J. (2011). Dissociating neuralcorrelates of action monitoring and metacognition of agency. *Journal of Cognitive Neuroscience, 23,* 3620–3636. doi:10.1162/jocn_a_00052

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Neuroscience, 2,* 661–670.

Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology, 86,* 838–848. doi:10.1037/0022–3514.86.6.838

Metaphor

The cognitive linguistic view of metaphor has a large number of implications for the social sciences. This entry discusses five general issues related to metaphor that are important for all social sciences: (1) the issue of the relationship between concepts and metaphors; (2) the issue of the types of understanding we have of the social world; (3) the issue of which metaphors are universal and which ones are nonuniversal, and why; (4) the issue of the distribution of metaphor in social-cultural life; and (5) the issue of the function of metaphor in social science.

Conceptual Metaphor Theory

It is useful to think of metaphor very broadly as conceptualizing one domain of experience in terms of another. The domain of experience that is used to comprehend another domain is typically more physical, more directly experienced, and better known than the domain we wish to comprehend, which is typically more abstract, less directly experienced, and less known. In the cognitive linguistic view of metaphor, originated in George Lakoff and Mark Johnson's *Metaphors We Live By*, the more concrete domain is called the *source domain* and the more abstract one is called the *target domain*. Domains of experience are represented in the mind as *concepts* given as mental frames. Hence, we talk about *conceptual* metaphors.

Thus, on this view, metaphor is a set of correspondences, or mappings, between the elements of two mental frames. For example, a set of correspondences between a traveler and a person leading a life, the way the traveler is going and the manner in which the person lives, the destination the traveler

wants to reach and the life goals of the person, the physical obstacles along the way and the difficulties the person has in life, all constitute the set of mappings that make up the conceptual metaphor *life is a journey*. A conceptual metaphor typically has a number of linguistic manifestations (metaphorically used words and more complex expressions) to talk about the target domain. In the example, the sentences "I *hit a roadblock*," "She *wanders aimlessly* in life," "This is not the *right way* to live," and so on, make manifest, or simply express, correspondences between the elements of obstacle and difficulty, destination and purpose, and path and manner, respectively. Taken together, they indicate that the highly abstract concept of life is understood in terms of the more concrete concept of journey.

There are several ways in which metaphors can be classified. They can be grouped according to their cognitive function, nature, conventionality, generality, grounding, and others. As regards *cognitive function*, conceptual metaphors can be structural (e.g., life is a journey) or nonstructural (as when, e.g., we evaluate a concept by assigning a positive or negative value to it—good is up, bad is down). As regards their *nature*, conceptual metaphors can be based on our general knowledge in connection with an area of experience or on the images we have of various domains of the world (e.g., the mind is a computer vs. the mind is a container). As regards *conventionality*, conceptual metaphors can be conventional or unconventional/novel (e.g., life is a journey vs. life is a box of chocolates). As regards *generality*, conceptual metaphors can be generic or specific (e.g., emotions are forces vs. anger is a hot fluid in a container vs. the angry person is a kettle). As regards *grounding*, conceptual metaphors may be grounded in analogical relationships between two domains or in bodily correlations in experience between the domains (e.g., life is a theater play vs. anger is heat). The kinds of metaphors these distinctions yield may combine with each other in particular cases of conceptual metaphors, and the distinctions come in various degrees between the two extremes.

Concepts and Metaphors

The subject matter of the social sciences is constituted by a large number of abstract concepts, such as emotions, politics, religion, society, nation, class, knowledge, communication, time, and so on. In the typical cases, concepts in general are characterized by a number of aspects. For this reason, since metaphors focus on a small number of aspects of concepts, abstract concepts are typically comprehended through a variety of different conceptual metaphors. It is not uncommon, however, that one or some of the conceptual metaphors pertaining to an abstract concept are more dominant than the others. Conceptual dominance can be defined and measured in terms of the productivity of a conceptual metaphor and/or the number of aspects of a target domain that it defines (by mapping source structure on it). This way, particular conceptual metaphors can emerge as "master metaphors" for particular abstract concepts. For example, as Zoltán Kövecses argues, the master metaphor for emotions is that emotions are forces, where, at least prototypically, emotion is viewed as a strong force overcoming a nonemotional self.

Folk Concepts and Expert Concepts

The *emotions-are-forces* metaphor reflects a naive, lay, or folk, understanding of emotions in Western cultures. Such naive, or lay, understandings of abstract concepts defined by our everyday metaphors are called *folk theories* of particular domains of experience. Folk theories are contrasted with *expert theories*, that is, theories that experts in a field propose, using their scientifically accepted methodologies. What is the relationship between a folk theory and an expert theory for a particular domain of experience? Sometimes, experts make unconscious use of folk theories, and they think about a particular domain that is closely related to a folk theory. A case in point is Freud's famous hydraulic model of the emotions, which is a version of the folk theory based on the *emotions-are-forces* conceptual metaphor, as found in everyday language in many Western languages. In many other cases, however, experts make a conscious effort to reconceptualize a domain. For example, the domain of time is widely conceptualized as space, yielding the *time-is-space* conceptual metaphor, but experts, such as philosophers, commonly argue that it is mistakenly conceptualized as space. It is, of course, also possible that particular expert theories become (part of) our folk theories.

Universality and Variation in Metaphor

Conceptual metaphors can be (near-) universal or language/culture specific. We get universal metaphors

when the source and target domains are grounded in universal bodily experience. Thus, the angry person as a pressurized container is a near-universal metaphor. With globalization, however, analogy-based cultural metaphors, such as *the mind is a computer*, can also become widespread, if not universal. Variation in metaphor can occur both cross-culturally and within cultures, as well as through time, as Kövecses has shown. The main causes of metaphor variation include differences in the key explanatory concepts of cultures and subcultures; physical environment; differential social and personal history; differential concerns of societies, subcultures, and persons; as well as differential uses of otherwise universal cognitive processes (such as metaphor and metonymy) within and across cultures.

Multimodality

Research by Charles Forceville has shown that conceptual metaphors also become visually manifest in cultures. Since conceptual metaphors are by definition in the mind, they can be widely distributed in culture, in the form of pictures, films, cartoons, and even material objects. Furthermore, our conceptual metaphors often define the actions people or organizations take or could take in society. George Lakoff argues that the conception of a nation in terms of *the nation is a strict-father family* versus *the nation is a nurturant-parent family* has far-reaching implications for political ideas and actions in American society.

Functions of Metaphor in Social Science

Metaphors have important functions in the social sciences. In order to understand a large number of concepts and issues important in the social sciences, folk theories are investigated. As we have seen, these are routinely constituted by metaphors. Furthermore, through the application of a master metaphor, conceptual metaphors can actually constitute a particular expert theory. Metaphor can also present a new way of understanding what understanding of the world is all about. Jacques Derrida's idea that the world is nothing but a text with an infinite number of interpretations can be regarded as a conceptual metaphor: *the world is a text*. Conceptual metaphors are also used for social criticism, especially in postmodernist theory. An example of a metaphor performing such a function is Julia Kristeva's theory of abjectness. We can interpret this idea as another conceptual metaphor, namely, *social abjectness is physical abjectness* or *social marginality is bodily marginality* (body fluids).

Zoltán Kövecses

See also Cognitive Sciences; Emotions; Language and Society; Language, Philosophy of; Postmodernism; Semantics and Pragmatics

Further Readings

Forceville, C. (2008). Metaphor in pictures and multimodal representations. In R. W. Gibbs Jr. (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 462–482). New York, NY: Cambridge University Press.

Gibbs, R. W., Jr. (Ed.). (2008). *The Cambridge handbook of metaphor and thought*. New York, NY: Cambridge University Press.

Kövecses, Z. (2000). *Metaphor and emotions*. New York, NY: Cambridge University Press.

Kövecses, Z. (2005). *Metaphor in culture*. New York, NY: Cambridge University Press.

Kövecses, Z. (2010). *Metaphor: A practical introduction* (2nd ed.). New York, NY: Oxford University Press.

Lakoff, G. (1995). *Moral politics*. Chicago, IL: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Metaphysics and Science

Metaphysics in the sense of the treatise that is known as Aristotle's *Metaphysics* is not concerned with entities that are supposed to exist beyond the empirical world but with the fundamental traits of the empirical world itself, taking into account the scientific theories about the world. That is why this treatise was ranged behind (or after—*meta* in Greek) the treatise on *Physics* and was called *Metaphysics*. This Aristotelian sense is the one that is pertinent to today's metaphysics, in contrast to an enquiry into entities that are supposed to exist beyond the empirical world (what poeple usually call "supernatural").

This entry first distinguishes between two different methodologies in contemporary metaphysics, then considers the relationship between metaphysics and science as regards fundamental science,

and finally focuses on functionalism as the main paradigm in the philosophy of the special sciences, including the social sciences.

Two Different Methodologies

There are two different methodological approaches in contemporary metaphysics—one that is based on a priori reasoning in the form of conceptual analysis ("armchair metaphysics") and another that sets out to naturalize metaphysics by seeing metaphysics as being continuous with science.

However, the contrast between these two approaches is not as sharp as it might seem at first glance. A priori reasoning in the form of conceptual analysis also depends on science mainly for the following two reasons: (1) The meaning of the concepts analyzed is subject to change, that change being notably due to scientific discoveries; thus, for instance, the meanings of concepts such as "particle," "atom," "soul," "mind," or "consciousness" are in more or less constant change as a result of progress in science. (2) The claims about the world put forward by metaphysicians subscribing to the method of conceptual analysis are based on empirical knowledge derived from science. For instance, the vast majority of metaphysicians today, including those subscribing to the method of conceptual analysis, defend a thesis of global supervenience according to which, to put the idea in simple terms, everything there is in the world is fixed by what there is in the fundamental physical domain. Thus, an exact physical duplicate of the real world would also contain all the people and their social institutions and interactions that there are in the real world. But this is a contingent claim, being based on modern science. It is maintained only with respect to possible worlds that are physically identical with the real world, not with respect to all possible worlds. By contrast, if one takes mathematics to be the paradigm of a priori knowledge, then one usually considers mathematics as applying to all possible worlds.

As regards naturalized metaphysics, there is no question of metaphysics being simply determined by science. If science were able to speak for itself, there would be no need for metaphysics at all. Scientific theories need interpretation in the sense of an answer to the question of what the world is like, provided that the scientific theory in question is true (or approximately true) and its interpretation cannot be simply read off from a scientific theory or settled by experiments. It requires philosophical argument. In fact, the actual practice of metaphysics is situated somewhere between the two methodological extremes of conceptual analysis and naturalized metaphysics.

The interplay between these two methodologies—conceptual analysis and naturalism—is also evident in the influential contemporary current known as *metaphysics of science*: Central topics considered by this current include the metaphysics of properties and laws, both in the fundamental and in the special sciences. Developing an answer to questions such as whether properties are purely qualitative (categorical) or dispositional (i.e., include the power to produce certain effects) and whether laws are contingent regularities or have a modal nature (necessity) requires both conceptual analysis as well as the consideration of scientific theories and explanations.

Metaphysics and Science in Physics

The fact that scientific theories call for an interpretation that requires philosophical argument is best illustrated by considering the current fundamental physical theory about matter, namely, *quantum mechanics*. The very formulation of the physical theory of quantum mechanics poses a problem that is known as the measurement problem. The problem is that the core formalism of quantum theory does not include the means to describe classical physical properties, that is, properties that have a definite numerical value, such as, for instance, either "plus" or "minus," and that occur notably at a definite location in space and time. Measurement outcomes, including the outcomes of measurements of quantum systems, are a paradigmatic example of classical properties. There are several strategies available to solve this problem, each of which leads to a different formalism of quantum mechanics, without there being a prospect in view of deciding between these strategies on the basis of experiments. That is why any solution to the measurement problem—and thus the very formulation of quantum mechanics—requires philosophical argument.

As regards space and time, since the days of Newton and Leibniz there has been a famous debate about whether space and time exist on their own, independent of the matter that is located in them

(substantivalism) or whether space and time consist only in the spatiotemporal relations between material entities (relationalism). Einstein's theory of general relativity has changed the features of this debate, but the question of the status of space and time is still hotly debated in today's metaphysics, and any answer to this question requires both taking the physics into account and developing a metaphysical argument.

Functionalism

Physics is a universal science, applying to the universe as a whole. Biology, psychology, economics, the social sciences, and so on, in contrast, are special sciences, since they apply only to a particular domain each. Functionalism is considered to be the appropriate paradigm as regards these sciences. Functionalism can be characterized in terms of the following three steps, quoted from Michael Esfeld and Christian Sachse (2011):

> (1) *The properties that constitute the domain of the special sciences are functional properties*. These are causal properties: they consist in having certain specific effects, given normal conditions, and possibly also in having certain specific causes.
>
> (2) *The functional properties that constitute the domain of the special sciences are realized by physical properties*. There is in each case a configuration of physical objects whose physical relations among themselves are such that, given normal conditions, they bring about qua configuration those effects that characterize a functional property of the special sciences. That is why they realize the functional property in question.
>
> (3) *The functional properties that constitute the domain of the special sciences can be multiply realized by physical properties*. Configurations of physical objects that are composed in different manners and that therefore come under different physical types (classifications) can nevertheless all realize a functional property of the same type of a special science, since they all bring about the same salient effects qua configurations under normal conditions. (p. 13)

Thus, for instance, the property of being a certain gene is a functional property, consisting in bringing about certain phenotypic effects under normal conditions (1). Genes are realized by certain configurations of molecules, notably certain DNA sequences (2). Molecular configurations (DNA sequences) that are composed in different physical manners can all produce qua configurations under normal conditions, those phenotypic effects that characterize a gene of a certain type. Consequently, genes can be *multiply realized* by molecular configurations of different types (3).

Steps 2 and 3 are the matter of empirical science, concerning the discovery of the realizers of a given functional property. As regards Step 1, again the methodological contrast between conceptual analysis and naturalism shows up: According to analytical functionalism, the functional definition of the properties that constitute the domain of the special sciences is a matter of conceptual analysis; according to scientific functionalism, such definitions have to be taken up from the most advanced scientific theories of the domain in question and are subject to change following scientific progress.

Even if one adopts the methodology of naturalism, functionalism leaves a number of issues open that depend on the scientific theories in question but that can only be settled by invoking philosophical, metaphysical argument. The most important of these issues are the following three interrelated ones:

1. *The status of the properties of the special sciences:* According to the position known as *role functionalism*, the properties with which the special sciences deal are second-order properties; having functional role properties consists in having certain other, first-order properties that realize the role in question; and functional properties thus are distinct from physical properties. According to the position known as *realizer functionalism*, there are only the realizer properties, in the sense that the functional properties with which the special sciences deal are identical with the physical realizer properties.
2. *The question of the reducibility of the special sciences:* The multiple realizability of functional properties excludes that the concepts of the special sciences that seize these properties can have the same extension as the concepts figuring in physical theories. Nonetheless, if a physical configuration of a certain type realizes

a functional property of a certain type, one can give a reductive explanation of why a functional property of the type in question occurs under certain circumstances. It is in dispute whether on this basis it is, in principle, possible to reduce the theories of the special sciences to physical theories despite multiple realization.

3. *The question of the causal efficacy of the properties of the special sciences:* According to what is known as the *completeness of the physical domain*, there is for any change in physical properties a sufficient physical cause (insofar as there is a cause at all). If the properties with which the special sciences deal are realized by physical properties, then any change in the domain of the special sciences requires a physical change. The question, therefore, is how the functional properties of the special sciences can be causally efficacious given the completeness of the physical domain. In particular, as noted by Jaegwon Kim, it is in dispute whether vindicating the causal efficacy of the properties of the special sciences requires regarding them as being identical with the physical properties that realize them.

Michael Esfeld

See also A Priori and A Posteriori; Analytic/Synthetic Distinction; Causation, Philosophical Views of; Laws of Nature; Reductionism in the Social Sciences; Supervenience

Further Readings

Albert, D. Z. (1992). *Quantum mechanics and experience*. Cambridge, MA: Harvard University Press.

Bird, A. (2007). *Nature's metaphysics: Laws and properties*. Oxford, England: Oxford University Press.

Earman, J. (1989). *World enough and space-time: Absolute versus relational theories of space and time*. Cambridge: MIT Press.

Esfeld, M., & Sachse, C. (2011). *Conservative reductionism*. New York, NY: Routledge.

Jackson, F. (1998). *From metaphysics to ethics: A defence of conceptual analysis*. Oxford, England: Oxford University Press.

Kim, J. (1998). *Mind in a physical world: An essay on the mind-body problem and mental causation*. Cambridge: MIT Press.

Ladyman, J., & Ross, D. (2007). *Every thing must go: Metaphysics naturalised*. Oxford, England: Oxford University Press.

Lewis, D. (1994). Lewis, David: Reduction of mind. In S. H. Guttenplan (Ed.), *A companion to the philosophy of mind* (pp. 412–431). Oxford, England: Blackwell.

Putnam, H. (1975). The nature of mental states. In H. Putnam (Ed.), *Mind, language and reality: Philosophical papers* (Vol. 2., pp. 429–440). Cambridge, England: Cambridge University Press.

METHODENSTREIT

The *Methodenstreit* (dispute over methods) began in 1883 as an argument between the Austrian school economist Carl Menger of the University of Vienna and the German Historical school scholar Gustav von Schmoller of the University of Berlin. The dispute ostensibly concerned the appropriate research methodology for economics—methodology being a fundamental concern for any area of study. Unfortunately, neither disputant seemed aware that their respective methodological arguments sprang from different theories of knowledge (epistemologies). Consequently, their contrasting views on the theories of concepts and of causality were never debated. Instead, there was a rancorous and inconclusive exchange of arguments on subsidiary questions, creating a lasting enmity between the two schools.

This entry summarizes the debate, identifies the key epistemological differences between the two disputants, and suggests that conflicts over research methodology may stem from deeper and less easily resolvable conflicts.

The German Historical School

The German Historical school is conventionally divided into the Older Historical school, the main figures of which were Wilhelm Roscher, Karl Knies, and Bruno Hildebrand, and the Younger Historical school, of which Gustav von Schmoller was the leading figure. The school began in the early 19th century partly as a reaction against the British Classical school of economics, the main figures of which were Adam Smith, Thomas Robert Malthus, and David Ricardo.

German scholars argued that the abstract deductive reasoning used by Classical school economic theorists created a separation between theory and

empirical reality. In addition, the Classicals used the individual person as their unit of analysis, rather than treating a social community as being itself an organism. The result was what Germans regarded as a universalist "dogma" of qualified laissez faire, and one that consequently supported the individualistic political philosophy that is current in Great Britain. In contrast, German historians argued that theoretical statements could only be empirical generalizations taken from actual experience and the history of a social organism itself. German experience being necessarily different from British, German institutions and economic laws would differ from those applicable to a British context. They argued that the German context was one of institutions and laws supporting German nationalism and the role of the state as an active force for social reform in Germany.

While members of the Older Historical school studied the economic development of nations using empirical generalization—a "historical method"—they differed among themselves concerning the question of the strictness of any generalizations, uniformities, or "laws" that they might find. They were united in the belief that there was no universal theory applicable to all cultures and throughout history.

Carl Menger's Critique and the Debate

Carl Menger began his 1883 published critique of the Historical school by asserting that the progress of economic science in Germany was blocked by an erroneous methodology—namely, the historical method. As a discipline, Menger divided economics into the areas of economic history, economic theory, economic policy, and public finance. He argued that general knowledge is only possible using economic theory, and no historical method—as the Historicists conceived it—could generate the theoretical laws needed to understand economic history, formulate policy, or practice public finance.

In his discussion of the derivation of the most important economic laws, those he termed the "exact laws" of economic theory, Menger made epistemological statements similar to those of contemporary Aristotelian "moderate realists." He did this in presenting both his theory of basic economic concepts and his theory of economic causality. He argued that economic concepts are derived from the essences or "simplest elements" of everything real, especially those of human nature and everyday life experience. The "exact" economic laws of economic theory are the result of logical deduction from those economic concepts. Menger's approach thus made the basic concepts and most important theoretical laws of economics not only ahistorical but also universal and an exemplification of intrinsic necessity—the ontology of the real.

Menger dismissed the idea of a "social organism," in favor of the concept of social phenomena as the collective result of the efforts of individual thinking and choosing human beings. He also argued that economic research methods must differ according to the formal knowledge of each area of research and that no one method is applicable to all areas in either the social or the natural sciences.

Schmoller responded to Menger's book with an adverse review in his journal, the principal one of the Historical school. He argued that Menger's concepts and abstract theorems not only lack reality but were also parochial and that his "atomistic" approach would miss the essence of social life. In contrast, the descriptive work of history and statistics was the foundation for economic theory. The more complete the description of a thing, the clearer would be the understanding of it and its relations to other things. Furthermore, "our law of thinking" requires that observed sequential events be recognized as revealing a current causal necessity. More careful and complete observations reveal more causal complexity in the connections of things and the necessary relativity of economic laws, tied as they are to the context of observation. Especially, individuals in society must be considered in their relationships to the mass. Schmoller's notion of causality resembles that of David Hume's nominalism, while his theory of concepts, although empirical, lacks the arbitrariness of a consistent nominalism. Whether nominalist or not, the contrast with Menger's epistemological views on concepts and causality is obvious. Menger responded in an abusive polemic, to which Schmoller replied by printing an insulting letter in his journal.

The polemics and inconclusive nature of the debate have led to negative assessments by almost all historians of economic thought. Recognizing that conflicting methods may imply epistemological conflict provides a basis for a reassessment.

Samuel Bostaph

See also Austrian Economics; Causation in the Social Sciences; Historicism; Holism, in the Social Sciences; Individualism, Methodological; Weber and Social Science: Methodological Precepts

Further Readings

Bostaph, S. (2005). The methodological debate between Carl Menger and the German historicists. In E. W. Younkins (Ed.), *Philosophers of capitalism* (pp. 113–131). Oxford, England: Lexington Books.

Mises, L. V. (1969). *The historical setting of the Austrian school of economics*. New Rochelle, NY: Arlington House.

Schulak, E. M., & Unterköefler, H. (2011). *The Austrian school of economics: A history of its ideas, ambassadors, and institutions* (A. Oost-Zinner, Trans.). Auburn, AL: Ludwig von Mises Institute.

Shionoya, Y. (Ed.). (2001). *The German historical school: The historical and ethical approach to economics*. London, England: Routledge.

MICROFOUNDATIONALISM

This entry analyzes the conceptual features and the wide use of the important concept of microfoundation in social ontology and social explanation.

Conceptual Analysis of Microfoundations

The concept of *microfoundations* is an important component within contemporary philosophy of social science. A microfoundation for a given compositional social phenomenon is a specification of the ways properties, structural features, and causal powers of a social entity are produced and reproduced by the actions and dispositions of socially situated individuals.

This concept is relevant to social ontology in this way. Social entities are understood to be *compositional*; they are assemblages constituted and maintained by the mentality and actions of individuals. So providing an account of the microfoundations of a structure or causal connection is a specification of the composition of the social-level fact. It is a description of the *agent-level* relationships and patterns of behavior that cohere in such a way as to bring about the higher-level structure or causal relationship.

The methodological requirement of microfoundationalism is a requirement regarding the validity of social explanations based on these assumptions about social ontology. It can be formulated in these terms: *An explanation of a macro-social phenomenon must be accompanied by a sketch of plausible microfoundations for the causal linkages it postulates*. In other words, if we assert a causal or explanatory relation between one social entity or condition and another, we must be prepared to offer a credible sketch of the ways in which this influence is conveyed through the mentalities and actions of individuals.

Much depends, however, on what precisely we mean to require of a satisfactory explanation: a full specification of the microfoundations in every case or a sketch of the way a given social-level process might readily be embodied in individual-level activities. If we accept the second version, we are licensing a fair amount of autonomy for the social-level explanation; whereas if we endorse the first version, we are leaning toward a requirement of reductionism from higher to lower levels in every case. It is most reasonable to interpret the requirement in the second way; it does not seem necessary to disaggregate every claim such as "Organizational deficiencies at the Bhopal chemical plant caused the devastating chemical spill" onto specific individual-level activities. We understand pretty well, in a generic way, what the microfoundations of organizations are, and it isn't necessary to provide a detailed account in order to have a satisfactory explanation.

The theory of microfoundationalism is very consistent with the methodology of *social mechanisms* as a basis for social explanations. When we ask about the microfoundations of a social process, we are asking about the mechanisms that exist at a lower level that create and maintain the social process.

Microfoundationalism falls in the general area *of methodological individualism* and *reductionism*, in that it insists on the compositional nature of the social. However, there is a recursive aspect of the theory that distinguishes it from strict reductionism. The individuals to whom microfoundations are traced are not asocial; rather, their psychology, beliefs, and motives are constituted and shaped by the social forces they and others constitute. So the microfoundational account of the workings of a social organization may well refer to the locally embodied effects of that organization on the current psychology of the members of the organization; their behavior in turn reproduces the organization in the next iteration. This is why some philosophers of social science, for example, David Little, prefer the idea of methodological localism over that of methodological individualism, while others, such as Peter Hedström, have preferred the idea of *structural individualism*.

The theory of *supervenience* is often invoked to express the idea that social entities and properties are constituted by individuals. Jaegwon Kim is the primary creator of the theory of supervenience in the philosophy of mind. This basic notion is expressed as the idea that a higher-level structure depends for its properties on the properties of its lower-level constituents. The advantage of the theory of supervenience is that it provides a way of recognizing the compositional nature of higher-level entities without presupposing explanatory reductionism from one level to the lower level. It is thus possible to allow for a degree of "explanatory autonomy" among factors that supervene upon lower-level processes if we are confident about the availability of microfoundations linking the supervening and lower levels.

Uses of "Microfoundations"

The explicit idea of microfoundations was first developed in the domain of *microeconomics*. There it referred to the necessity of deriving macroeconomic phenomena from the premises of rational economic behavior. Maarten Jansen describes the theory of microfoundations within economics as the attempt to understand aggregate economic phenomena in terms of the behavior of individual economic actors and the interactions among them.

The idea of microfoundations is now important in many areas of the social sciences, including especially sociology and political science. Early advocates of the microfoundations principle included contributors to the theories of analytical Marxism in the 1980s. David Little (1994) summarized the analytical Marxism approach in these terms:

> Marxist thinkers have argued that macro-explanations stand in need of microfoundations: detailed accounts of the pathways by which macro-level social patterns come about. These theorists have held that it is necessary to provide an account of the circumstances of individual choice and action that give rise to aggregate patterns if macro-explanations are to be adequate. (p. 479)

Particularly important within recent sociology were the theories formulated by James Coleman. Coleman does not use the term *microfoundations* explicitly, but his analysis of the relationship between the macro and the micro implies the requirement of providing microfoundations as a condition for good explanations in the social sciences. The "Coleman boat" linking macro and micro levels of explanation is a graphical way of representing the microfoundations of a macro-level fact.

The idea of microfoundations is also a core constituent of the methodology of analytical sociology. Peter Hedström, the leading advocate of the approach, does not use the term explicitly in *Dissecting the Social*, but the concept pervades the volume and the approach. Analytical sociology is interested in explaining macro-level social phenomena on the basis of social mechanisms at the level of the purposive agents who constitute social groups and processes.

The theory of microfoundationalism is often associated with the use of rational choice theory as a basis for explaining social outcomes. This association is not a necessary one, however. It is possible to maintain that agents are motivated by considerations that fall outside traditional rational choice assumptions without surrendering the idea that macro-level outcomes depend on the actions and thoughts of the individuals who make them up.

Daniel Little

See also Analytical Marxism; Analytical Sociology and Social Mechanisms; Individualism, Methodological; Holism, in the Social Sciences; Mechanism and Mechanismic Explanation; Reductionism in the Social Sciences; Supervenience

Further Readings

Coleman, J. S. (1990). *Foundations of social theory*. Cambridge, MA: Harvard University Press.

Hedström, P. (2005). *Dissecting the social: On the principles of analytical sociology*. Cambridge, England: Cambridge University Press.

Hedström, P., & Bearman, P. (Eds.). (2009). *The Oxford handbook of analytical sociology*. Oxford, England: Oxford University Press.

Hedström, P., & Swedberg, R. (Eds.). (1998). *Social mechanisms: An analytical approach to social theory*. Cambridge, England: Cambridge University Press.

Jansen, M. C. W. (1993). *Microfoundations: A critical inquiry*. London, England: Routledge.

Janssen, M. C. W. (2008). Microfoundations. In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave dictionary of economics*. London, England: Palgrave-Macmillan.

Kim, J. (1993). *Supervenience and mind: Selected philosophical essays*. Cambridge, England: Cambridge University Press.
Little, D. (1994). Microfoundations of Marxism. In M. Martin & L. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 479–496). Cambridge: MIT Press.
Little, D. (1998). *Microfoundations, method, and causation: On the philosophy of the social sciences*. New Brunswick, NJ: Transaction.
Little, D. (2006). Levels of the social. In S. P. Turner & M. W. Risjord (Eds.), *Handbook for philosophy of anthropology and sociology* (pp. 343–371). Amsterdam, Netherlands: Elsevier.
Roemer, J. (1981). *Analytical foundations of Marxism*. New York, NY: Cambridge University Press.
Udehn, L. (2001). *Methodological individualism: Background, history and meaning*. London, England: Routledge.
Weintraub, E. (1979). *Microfoundations*. Cambridge, England: Cambridge University Press.
Zahle, J. (2007). Holism and supervenience. In S. P. Turner & M. W. Risjord (Eds.), *Philosophy of anthropology and sociology* (pp. 311–341). Amsterdam, Netherlands: Elesevier.

Mill and the Moral Sciences

John Stuart Mill (1806–1873) wrote extensively on the philosophy of the moral sciences, and in many key respects, he remains one of the leading champions for positioning the moral sciences on an epistemic par with the natural sciences. Mill's *oeuvre* fills an entire library shelf, but he is mostly known for two short works, *On Liberty* (1859) and *Utilitarianism* (1861), both of which remain pillars of the philosophical canon, especially as far as moral and political philosophy is concerned. But his philosophical assessment of the moral sciences can best be found in his *System of Logic* (1843), *Principles of Political Economy* (1848), *Considerations on Representative Government* (1861), *Examination of Sir William Hamilton's Philosophy* (1872), and various essays, particularly "On the Definition of Political Economy and on the Method of Investigation Proper to It" (1836) and "On Comte and Positivism" (1865).

As an empiricist, Mill believed that there are only particulars in the world. Nevertheless, there are sufficient uniformities among groupings or types of particulars such that one can arrive at laws, synchronous or successive. Mill recognized that there were many types of laws, notably conservation laws, causal laws, universal laws, tendency laws, as well as laws about emergent properties, or chemical laws. He much appreciated the tendency toward the unification of laws, particularly Newton's synthesis of the known laws of mechanics and astronomy. More significantly, our belief in the overarching uniformity of nature is deepened over time as each newly discovered law adds inductive support.

Although Mill admired the work of August Comte, he did not adopt Comte's hierarchical and reductionist schema for the sciences. Mill believed that there were several distinct moral sciences and each one had its own salient phenomena and set of methods. Following a number of the Scottish Enlightenment philosophers, including his own father, Mill viewed the study of the mind as the most fundamental moral science. He endorsed the main psychological theory of his day, known as *associationism*. Our specific perceptions yield impressions which then form ideas; these ideas can be joined with other ideas in accordance with three laws of association: resemblance, contiguity, and causation. Hence, the mechanisms of the human mind by which ideas are processed are uniform, and hence, there are laws that govern the formation of beliefs and actions. For Mill, psychology was a strongly empirical science that sought correlations between physical states of the brain and the body with inner states of the mind. Introspection was a key asset in these investigations.

Of all the moral sciences, Mill left his mark on political economy. He first cut his teeth on the work of the classical political economist David Ricardo at the age of 13, and much later, he sought to reinstate the leading Ricardian tenets embedded in a broader social and political context. Political economy was defined as the science that sought the laws of the production and distribution of wealth. It used the same deductive method as physics. It commenced with inductive generalizations drawn both from psychology, for example, that all humans seek to obtain the greatest amount of wealth with the minimum amount of exertion, and from the physical sciences, for example, the chemical laws that govern fermentation. From this are derived the laws that govern the production and distribution of wealth, for example, the inverse relation between profits and wages, or the principle of comparative advantage.

From these laws are derived specific observations, which are then matched to the world for verification. Since the initial laws require abstractions and idealizations, Mill maintained that the third step of verification also involved the search for disturbing causes, which would in turn draw upon other psychological, historical, or cultural dispositions, that is, patterns from the other moral sciences.

Mill also believed that there was great potential for the sciences of sociology and political science. The former was distinctively historical in his view, and therefore the laws would be located in the succession of social states. The closing chapters of his text on political economy offer a sketch of his sociology. Political science, by contrast, was conceived of as a deductive science using methods similar to economics. Although Mill laid out his own distinct system of representative government, he failed to ascend to the theoretical level.

Mill also believed that human nature evolves over time and that many of the laws for the moral sciences will thus be historically contingent. To attend to cultural and physical circumstances, Mill proposed a new science that he called "ethology," the study of human character. Although it held a central position in his schema for the moral sciences, Mill never followed up on this proposal; nor did he find disciplines, and as a result, ethology was stillborn. Nevertheless, it served as a reminder that for Mill, human nature was not just a subject for scientific study but could be significantly improved by absorbing scientific discoveries.

Margaret Schabas

See also Deduction; Empiricism; Individualism, Methodological; Intention, Social Psychology of; Laws of Nature; Philosophy of Economics, History of; Positivism, History of; Reductionism in the Social Sciences; Scottish Enlightenment: Influence on the Social Sciences

Further Readings

Hausman, D. (1981). John Stuart Mill's philosophy of economics. *Philosophy of Science, 48,* 363–385.

Hollander, S. (1985). *The economics of John Stuart Mill* (2 vols.). Toronto, Ontario, Canada: University of Toronto Press.

Mill, J. S. (1962–1991). *Collected works of John Stuart Mill* (33 vols.; J. M. Robson, Ed.). Toronto, Ontario, Canada: University of Toronto Press. (Original work published 1812–1873)

Schabas, M. (2005). *The natural origins of economics*. Chicago, IL: University of Chicago Press.

Skorupski, J. (Ed.). (1998). *The Cambridge companion to Mill*. Cambridge, England: Cambridge University Press.

Wilson, F. (1990). *Psychological analysis and the philosophy of John Stuart Mill*. Toronto, Ontario, Canada: University of Toronto Press.

Mind–Body Relation

After the transgression in the Garden of Eden, which sealed our fates, God told Adam, "Dust thou art, and unto dust thou shalt return" (Genesis 3:19). *Genesis* tells us that God fashioned Adam from the dust of the earth and Eve from Adam's rib. The Old Testament contains no hint that we might have an immaterial soul. Nor is there any clear evidence for the presence of that idea in the New Testament. Eternal life for us will come, the Bible tells us, through the resurrection of the body. Several 100 years after the writing of *Genesis*, and a few 100 years before the birth of Jesus, Socrates sat at peace in his jail cell, surrounded by his students, awaiting the hemlock he was ordered by the court to drink as punishment for corrupting the young of Athens (Plato, 1961, pp. 40–98). His students marveled at his calm and wondered why he didn't fear his impending death. The reason, it turns out, is that he didn't really think that he was going to die but only that his body was going to die. He maintained that he had a soul that was not made of material parts, and so could not come apart and thereby cease to exist. The idea that we have immaterial souls seems to have entered Christian theology through the work of Augustine of Hippo, an African philosopher who tried to integrate Platonic philosophy with the teachings of Christianity (Van Inwagen, 1975). As concerns the question how our immaterial souls are related to our material bodies, Augustine tells us, "The manner in which spirits are united to bodies is altogether wonderful and transcends the understanding of men" (cited in Haldane, 1994, p. 335).

Dualism

By the 17th century, it was well-known among the intellectual class that the soul or mind is related in some intimate way to the brain. (That was not known in antiquity. Aristotle had no idea that the

soul or mind is related to the brain.) Rene Descartes (Cottingham. Stootoff, & Murdoch, 1985) maintained that the relationship between the two is one of cause and effect and speculated that the immediate locus of causal interaction is in the pineal gland of the brain. States of and changes in the mind can directly cause, and can be directly caused by, states of and changes in the pineal gland. There is mental-to-physical causation when we perform actions. Our minds produce changes in the physical realm by producing changes in the pineal gland, changes that cause other brain events that lead to contractions of muscles and, thereby, to movements of or postures in our skeletons or vocal cords, which, in turn, issue in environmental effects. There is physical-to-mental causation when we engage in sense perception: when we see, hear, smell, taste, or feel something. For example, when we see, the scene before our eyes directs light to our eyes, which initiates a causal chain of events that ultimately results in our having a visual experience of the scene—that is, with the scene's looking some way to us. This doctrine of two-way causal interaction is called "Cartesian Interactionism." Descartes also held that although the mind is located in time, it is not located in space. Minds are nonspatial substances (whose essence is to think) and thus are distinct from any physical substance since all such substances are in space. This doctrine is called "Cartesian Substance Dualism."

Princess Elizabeth of Bohemia saw a tension between this substance dualism and causal interactionism, one that was especially acute since at the time of her correspondence with Descartes it was held that an object can directly causally influence the motion of another object only if the objects are in spatial contact. She wrote to Descartes asking

> how the mind of a human being can determine the bodily spirits in producing voluntary actions . . . For it appears that all determination of movement is produced by the pushing of the thing being moved, by the manner in which it is pushed by that which moves it, or else by the qualification and figure of the latter. Contact is required for the first two conditions, and extension [in space] for the third. [But] you entirely exclude the latter from the notion you have of the soul, and the former seems incompatible with an immaterial thing. (Cited in Garber, 2001, p. 172)

In response to this how-question, Descartes maintains that although contact is required for direct physical causation, the mind and body are united in such a way that contact is not required for them to directly causally interact. His notion of mind–body union, however, went unexplained, and Princess Elizabeth was unconvinced. She responded, "And I admit it would be easier for me to concede matter and extension to the mind than it would be for me to concede the capacity to move a body and be moved by a body to an immaterial thing" (cited in Garber, 2001, p. 134). Even without the assumption that direct causation requires spatial contact (an idea that Newton eschewed for gravitational influence), it is indeed deeply mysterious how states of and changes in something not located at all in space can cause or be caused by states of and changes in something in space. Such causal transactions would have to be brute facts, facts that admit of no explanation; for there cannot be a mechanism that explains how such transactions occur, and so there can be no answer to the how-causation. (For a discussion of further difficulties, see Kim, 2006, chap. 2.) Descartes himself recognized the mystery. He made a remark at one point that is reminiscent of Augustine's remark quoted above: "It does not seem to me that the human mind is capable of conceiving quite distinctly and at the same time both the distinction between mind and body and their union" (cited in Kenny, 1970, p. 142).

Early Reactions to Dualism

In response to Descartes, Thomas Hobbes denied substance dualism, insisting that the mind is the brain. Gottlieb Leibniz and Nichols Malebranche denied interactionism, and both held versions of parallelism. Leibniz espoused the parallelist doctrine of preestablished harmony: There is no causal interaction between the mental realm and the physical realm, but there are mental–physical correlations because, of the infinity of possible worlds, God made actual a possible world in which such correlations hold. Malebranche espoused the doctrine of occasionalism: God is the only causal agent; he creates the world anew at each instant of time, but in a way that ensures mental and physical correlations. Baruch Spinoza held that God is the only substance and that the mental and the physical are just distinct modes or attributes of that one substance. He combined a kind of pantheism with a "dual-aspect theory of the mental and physical," a theory according to which mental and physical attributes are distinct attributes that are possessed by the same substance. Bishop Berkeley and Georg Wilhelm

Friedrich Hegel, among many others, held versions of idealism: the doctrine that everything is mental. According to Berkeley, everything is mental because everything is an idea in the mind of God.

Modern Science

If contemporary science can be trusted, *Genesis* is correct: We are made of the dust of the Earth—kinds of dust that science tells us were formed inside stars. We are thus made up of star dust. But there is no mind dust. The dust from which we are composed can be found not only in things that lack minds but even in nonliving things, such as rocks and water. To be sure, we are made up of cells, which are organic matter. But cells are made up of molecules, which are made up of atoms, which are made up of electrons, protons, and (with the exception of the hydrogen atom) neutrons; and neutrons and protons are in turn made up of quarks. Perhaps at the bottom of this constitutional hierarchy are not physical particles but rather something physical that is not particle-like: strings of energy or m-branes. Or perhaps there is no bottom level. It could be that below every peel of the physical onion is another peel. But be that as it may, we are composed of electrons, neutrons, and protons, just as are rocks. What differences there are between rocks and human brains are due to differences in the number and complexity of organization of these constituents. Moreover, it seems that the fundamental forces at work in sustaining the organizations and producing changes in organization, no matter how complex the organization, are all forces that are exerted below the level of complexity of atoms. Mechanics, a branch of physics, aims to be a comprehensive theory of motion: to formulate the general laws that all motion obeys (in Newtonian mechanics, they are the three laws of motion; in nonrelativistic quantum mechanics, the law is Schrödinger's equation) and to describe the fundamental causes of motion (in Newtonian mechanics, the force laws; in quantum mechanics, the Hamiltonians for the Schrödinger's equation). Mechanics tells us that the fundamental forces, the ones that are the bases of all forces, are gravity, electromagnetic force, weak force, and strong force. These forces are all ones that are exerted below the level of complexity of atoms. Electromagnetic force and weak force may be one force, and physicists hope for further reduction in the number of fundamental forces. But the idea often espoused in the history of science that there are fundamental forces that exert their influence only when matter becomes configured in certain complex ways is not incoherent. Contemporary mechanics, however, thus far has no need of that hypothesis; and the current received view is that there are no such forces. Moreover, even if physics were to discover that there is such a fundamental emergent force, physics would take the force to be due to microstructural properties of the relevant complex configurations of matter, not to an immaterial object (McLaughlin, 1992).

It should be noted, however, that science itself does not tell us that we have no immaterial part. (According to general relativity theory, though, nothing can be in time without being in space, even if only at a point in space.) It is, rather, that science has no need of that hypothesis in its aim to provide a comprehensive theory of motion. Our bodies and its organs at least, including our brains, are complex physical objects wholly made up of atoms and more fundamental physical particles. Moreover, the movements of our bodies and of its parts are governed by the blind physical forces of nature. How can we be constituted by a complex physical object, the movements of which and the movements of the parts of which are governed by the blind physical forces of nature and yet have a mind?

It has been argued that a mind isn't a special sort of object, material or immaterial (Ryle, 1949). What it is to have a mind is just to have certain kinds of abilities and capacities (abilities to acquire abilities). These abilities include in our case the ability to think, to sense, and to feel. When our brains are functioning properly, they are the material basis of our mental abilities. But how could such abilities be exercised in the brain? The exercise of such abilities will involve mental states that participate in mental processes. How could there be such states and processes within a brain?

Epiphenomenalism

In the last third of the 19th century, but even prior to the discovery of neurons in 1890, some biologists realized that the search for gaps in neurophysiological processes in the brain that might be filled by mental states was futile. The absence of such gaps led Thomas Huxley (1874) to maintain that animals

are "conscious automata," by which he meant that they are physical machines that are also conscious but that conscious states are merely effects of the brain, never causes; they are causally inert. Using an old medical term for symptoms of diseases, William James (1890) claimed that this view treats mental phenomena as "epiphenomena." James Ward (1896) coined the term *epiphenomenalism* for this view. Epiphenomenalism is the view that mental states and events are caused by physical states and events but mental states and events never cause anything. On this view, feeling pain never causes one to wince or to withdraw one's limb or to cry out; indeed, feeling pain never causes one even to believe that one is in pain. Moreover, thinking is not a causal process in which one thought leads to another. There are no mental causal processes. If epiphenomenalism is true, we don't exercise mental abilities. There is only the appearance that we do, because we undergo certain types of temporally ordered patterns of mental states. Instances of types of mental states in such patterns do not causally interact; the appearance of causal interaction is due to the fact that pairs of mental states are dual effects of some common physical cause in the brain. Moreover, the experience of instances of such patterns does not cause us to believe that there are mental casual processes. Experiences have no causal effects. The conscious mind, an effect of the physical brain produced anew on each occasion of consciousness (think here of an analogy with occasionalism), is causally impotent, powerless to effect any change in the course of events but pathetically under the illusion that it can.

The "How-Question"

Suppose—as seems at the very least enormously plausible—that minds really can change the world through the exercise of a subject's mental abilities. When we exercise mental abilities, the mental states we are in participate in processes as causes. This takes us back, then, to the question of how a physical object, no matter how complex, could be in a mental state. We've raised a number of how-questions. But let's call this last one "the how-question."

One answer to the how-question is that mental abilities are just complex behavioral abilities and that mental states are just complex dispositions to behave. Such behavior includes verbal behavior, which is sufficiently complex in our case to set it apart from the kind of auditory signal systems used by other animals. Analytical behaviorists attempted to define mental terms (e.g., *believes*, *desires*, *intends*, *pain*, *itch*, etc.) in terms of dispositions to overt or peripheral behavior—behavior that involves the movement or posture of some part of the periphery of the body (Carnap, 1932/1933). But they failed to provide even a single example of such a definition that had any credibility. And it became clear why the attempts failed. What one is disposed to do doesn't depend only on a belief or a desire, and so on, but on a vast range of one's beliefs and desires, as well as one's intentions, hopes, wishes, fears, and other mental states. If asked whether one believes that it will rain tomorrow, whether one utters the sound /yes/ in response will depend on much more than whether one believes it will rain tomorrow; it will depend on whether one understands what is being asked, knows what the sound means, and thinks the asker will hear and on whether one wants to answer truthfully more that one wants other relevant things, a matter that will depend, in part, on other beliefs one has (Chisholm, 1957, chap. 2). It was also convincingly argued that even terms for bodily sensations, such as "pain," cannot be defined in terms of dispositions to behave (Putnam, 1968). It is now widely acknowledged that analytical behaviorism is mistaken, since such definitions are impossible.

There is, however, a progeny of analytical behaviorism that has adherents, namely Real-Patterns Intentional System Theory (Dennett, 1991, 1996). This view does not entail that mental terms can be (finitely) defined in terms of dispositions to peripheral behavior. Indeed, on this view, no mental term can be so defined. The holism of the mental is acknowledged. But it is claimed that it is nevertheless the case that when a mental term correctly applies to an individual, it does so by virtue of the individual's having an enormously complex global pattern of dispositions to peripheral behavior. On this view, there can be no difference between two individuals with respect to whether they have mental states without a difference in their global patterns of dispositions to peripheral behavior. A thesis of the form "There can be no A-difference without a B-difference" is called "a supervenience thesis," and if a thesis of this form is true, then A-respects are said to "supervene on" B-respects. According

to Real-Patterns Intentional System Theory, having mental states supervenes on global patterns of dispositions to peripheral behavior. The supervenience thesis is taken to be true because mental states *are* patterns of dispositions to peripheral behavior.

Functionalism

A different answer to the how-question is that to be in a mental state is just to be in a type of state that has a complex causal role, a role as cause and as effect, which includes causes and effects among certain internal states of the individual. This view is called "functionalism." It comes in a number of varieties. There are at least three important distinctions that cross-classify functionalist theories of mind.

The first distinction is between role functionalism and filler functionalism (which is sometimes called "realization functionalism"). According to role functionalism, a type of mental state is a second-order state, a state of being in some state of a certain sort, one with a certain kind of conditional causal role. On this view, a type of mental state is a state of being in some state type or other whose instances are caused in certain ways under certain conditions and whose instances have certain kinds of effects under certain conditions, where the causes and effects include environmental stimuli, behavior, and internal states (see, e.g., Loar, 1980). The first-order states that occupy the causal roles are said to "realize" the mental state; thus, realization is the relation of role occupancy; and it is possible for a type of mental state to be multiply realizable—that is, to be realizable by a variety of different kinds of first-order states. According to filler functionalism, a type of mental state is the state that has a certain causal role (as characterized above) or the state that has it in normal members of the species of animal in question (human or otherwise) or in beings with a certain kind of physical structure (Armstrong, 1968; Lewis, 1966; Smart, 1959). On the filler functionalist view, there is multiple realization only in the sense that the state that occupies the causal role in one kind of animal or structural being can be different from the state that occupies it in another.

The second distinction is between narrow role functionalism and wide role functionalism. A functionalism (role or filler) is narrow if the causal role includes not only proximal environmental causes but also distal ones. Wide role functionalism can be wide indeed. On some wide role-functionalist theories, evolutionary history is even included among the causes.

The third distinction is between analytical functionalism and scientific functionalism. According to analytical functionalism, the relevant causal roles are only those that common sense or "folk psychology" associates with the state type in question. According to scientific functionalism, which is often labeled "psycho-functionalism," the relevant causal roles are those that the correct scientific psychology will associate with types of mental states. This view is often associated with computational psychology, according to which the mind is a computer and states of mind are types of computational states of the computer (Fodor, 1975; Putnam, 1973; Turing, 1950).

All versions of functionalism are committed to the supervenience thesis that there is no mental difference between two individuals without a difference in the causal roles (wide or narrow) of the individual's internal states. The supervenience thesis is claimed to be true because a type of mental state is a type of state with a certain conditional causal role (filler functionalism) or is a state of being in some type of state or other with a certain kind of conditional causal role (role functionalism).

Anomalous Monism

Proponents of the doctrine of anomalous monism (Davidson, 1970) maintain that types of intentional mental states such as believing that p, desiring that p, intending that one A, hoping that p, fearing that p, and the like—so-called propositional attitudes—cannot be reduced to types of physical or functional states, because our propositional attitude concepts are governed by the normative ideal of rationality, while our physical and functional concepts are governed by very different constitutive principles. If anomalous monism is correct, then we can never get a satisfactory answer to the how-question. Nevertheless, anomalous monists maintain that every concrete occurrence of a propositional attitude or a change in one is identical with some physical state or event. They maintain that on the grounds that if events are causally related, then they fall under some strict law of nature and all such laws are physical laws.

Consciousness and Qualia

A number of contemporary philosophers who are materialists nevertheless maintain that types of states of consciousness cannot be reduced to

behavioral or functional states of any sort (Block, 1978; Searle, 1992). Even some philosophers who maintain that propositional attitudes are functional states of some sort deny that such reduction is possible for consciousness states (Block, 1978; Chalmers, 1996; Levine, 2003). These philosophers agree with Thomas Nagel's (1974) assessment: "Consciousness is what makes the mind-body problem really intractable." The term *consciousness* has a number of uses. But in the intended sense here, states of consciousness are states such that it is like something for the subject of the state to be in the state. Such states are called "states of phenomenal consciousness." They are experiences. It is, for instance, like something for a subject feeling pain to feel pain; and it is like something for the subject of a visual experience of red to visually experience redness. The what-it-is-like for a subject's aspect of a state of phenomenal consciousness is sometimes called its "phenomenal (or qualitative) character"; the term *qualia* is also often used for the what-it-is-like aspect. Let's use the term *qualia* (singular "quale"). These philosophers deny that qualia are patterns of dispositions to peripheral behavior or just a matter of causal-role properties of any sort.

In response, there have been rigorous, detailed attempts to show how qualia can be understood in wide-functional-role terms (see especially Dretske, 1995; Hill, 2009; Lycan, 1996; Tye, 2000). But there are at least six main, interestingly different views that both eschew any kind of behavioral or functionalist accounts of states of phenomenal consciousness and have defenders in the contemporary philosophical community. All the views have historical roots; some have roots tracing back to antiquity.

Emergent materialism (to use C. D. Broad's, 1925, apt name for the doctrine) is the view that qualia are fundamental, irreducible properties of certain physical states of the brain (neural states, say). The linkage between kinds of qualia and kinds of physical states of the brain is one of natural law: There are fundamental laws to the effect that if the brain is in a certain kind of physical state, then there is a quale of a certain kind. Emergent materialism is a property-dualist view that eschews substance dualism. It is a kind of dual-aspect view where the two aspects are aspects of the brain and are linked by fundamental laws of nature. Since the linkage is via fundamental laws of nature and since laws of nature are logically contingent, emergent materialists are committed to the logical possibility of what are nowadays called "zombies" (Chalmers, 1996), that is, beings that are exact physical duplicates of conscious beings but that are not themselves conscious. On a different version of emergentism, zombies are impossible in that they are not even coherently conceivable, but the specific characters of qualia are linked to patterns of physical states of the brain only by fundamental laws of nature (Kim, 2005). Both versions are thus committed to the view that there are fundamental laws of nature linking qualia with physical states. (Herbert Feigl, 1958, famously criticized this view, calling the laws in question "nomological danglers.")

Pansychism is the view that everything in space-time has qualia—literally everything, and so electrons, rocks, puddles of water, clouds, and galaxies (Drake, 1925; James, 1890; Rosenberg, 2005; Seager, 2009; Skrbina, 2005; Strawson, 1994, 2006, 2008). This is property dualism all the way up and all the way down the micro–macro scale and all the way up and all the way down the scale of levels of complexity of physical organization. (These are different scales. Brains are middle-sized objects, objects that are in the middle of the micro–macro scale. But they are the most complex objects known.)

What we may call "pan-protophyicism" is the view that qualia are reducible, but reducible to properties that we have not yet discovered and that are very different both from causal-role properties and from any properties that have been posited by physical science (Nagel, 1974, 1986, 1999). These unknown properties are labeled "proto-mental properties." The claim is thus that qualia reduce to proto-mental properties. A completed physics will need to invoke such proto-mental properties, but in doing so, it will not be a recognizable descendant of current physics. There is no telling now exactly how different it will be, but it could turn out to be as different from current physics as air–earth–fire–water chemistry is from current chemistry.

Mysterianism is the view that qualia are reductively explainable but that we human beings are cognitively closed to the properties that reductively explain them (McGinn, 1989). (Recall Augustine's claim and the quote from Descartes.) Just as a dog, for instance, can't understand the mathematical property of being a square root, we can't understand the properties that reductively explain qualia. Those properties run against the grain of our thought; we can't get concepts around them; they are beyond our comprehension. Mysterianism is compatible with pan-protophysicism, but the latter doesn't entail the

former. A pan-protophyscist typically maintains that human beings may someday (with work and luck) discover the proto-mental properties that reductively explain qualia.

So-called Russellian physicalism (because it was a view once held by Bertrand Russell, 1927) is the view that physical properties are dispositional properties that have intrinsic, qualitative, categorical bases, that the qualia we know by direct acquaintance are the intrinsic, qualitative, categorical bases of certain physical properties that are exemplified in our brains but that we cannot know the qualitative bases of other physical properties by direct acquaintance; indeed, we can know other physical properties only by theoretical description, and so these qualitative bases are inaccessible to us (see, e.g., Stubenberg, 1988).

Finally, type materialism for phenomenal consciousness is the view that qualia are identical with neurobiological properties of the brain, ones of a sort that could be posited by a recognizable descendant of current neurobiology (Hill, 1991; Hill & McLaughlin, 1999; Loar, 1997; McLaughlin, 2001, 2006, 2010, 2012; Papineau, 2002). Claims asserting such identities will be a posteriori, not a priori. The reason is that our concepts of qualia (e.g., the concept of the feeling of pain) are very different from neurobiological concepts, which are theoretical concepts; the concepts play very different roles in our cognitive economy. It will thus always be coherently conceivable that something could have a quale and not have neurobiological properties, and conversely for any neurobiological property. On this view, zombies are coherently conceivable; but it is maintained that they are nevertheless impossible. The properties to which our qualia concepts refer are neurobiological properties. We can determine the relevant identities only by empirically searching for the strict neurobiological correlates of qualia. The emergent materialist would claim that the correlations are fundamental laws of nature. The type materialist claims that, on the contrary, if we find such strict lawful correlations, then we could justify the claim that qualia are identical with their neurobiological correlates by inference to the best explanation of those correlations.

Each of the above positions faces a host of truly formidable difficulties, difficulties beyond the scope of this essay. Let it suffice to note that the place of phenomenal consciousness in nature remains a topic of intense interest and debate.

Brian P. McLaughlin

See also Behaviorism, Philosophical Conception of; Behaviorism in Psychological Explanation; Classical Computationalism, Connectionism, and Computational Neuroscience; Cognitive Sciences; Consciousness; Embodied Cognition; Idealism; Modularity of the Mind; Supervenience

Further Readings

Armstrong, D. M. (1968). *A materialist theory of mind.* London, England: Routledge & Kegan Paul.

Block, N. (1978). Troubles with functionalism. In C. W. Savage (Ed.), *Perception and cognition* (pp. 261–325). Minneapolis: University of Minnesota Press.

Broad, C. D. (1925). *The mind and its place in nature.* London, England: Routledge.

Carnap, R. (1932/1933). Psychology in physical language (G. Schick, Trans.). In A. J. Ayer (Ed.), *Logical positivism* (pp. 165–198). New York, NY: Free Press. (Reprinted 1959)

Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory.* New York, NY: Oxford University Press.

Chalmers, D. (2002). Consciousness and its place in nature. In D. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 247–272). New York, NY: Oxford University Press.

Chisholm, R. (1957). *Perceiving: A philosophical study.* Ithaca, NY: Cornell University Press.

Cottingham, J., Stootoff, R., & Murdoch, D. (Eds.). (1985). *The philosophical writings of Rene Descartes* (Vols. 1 & 2). Cambridge, England: Cambridge University Press.

Dennett, D. (1991). Real patterns. *Journal of Philosophy, 88,* 27–51.

Dennett, D. (1996). *The intentional stance.* Cambrdige: MIT Press.

Drake, D. (1925). *Mind and its place in nature.* New York, NY: Macmillan.

Dretske, F. (1995). *Naturalizing the mind.* Cambridge: MIT Press.

Feigl, H. (1958). The "mental" and the "physical." In H. Feigl, M. Scriven, & G. Maxell (Eds.), *Concepts, theories, and the mind-body relation* (pp. 370–497). Minneapolis: University of Minnesota Press.

Fodor, J. (1975). *The language of thought.* New York, NY: Thomas Y. Crowell.

Garber, D. (2001). *Descartes embodied*. Cambridge, England: Cambridge University Press.

Haldane, J. (1994). Medieval and Renaissance philosophy of mind. In S. Guttenplan (Ed.), *A companion to philosophy of mind* (pp. 333–337). Oxford, England: Oxford University Press.

Hill, C. S. (1991). *Sensations: A defense of type materialism*. Cambridge, England: Cambridge University Press.

Hill, C. S. (2009). *Consciousness*. Cambridge, England: Cambridge University Press.

Hill, C. S., & McLaughlin, B. P. (1999). There are fewer things than are dreamt of in Chalmers's philosophy. *Philosophy and Phenomenological Research, LIX,* 445–454.

Huxely, T. (1874). On the hypothesis that animals are automata, and its history. *Forthnightly Review, 22,* 555–580.

James, W. (1890). *The principles of psychology.* Cambridge, MA: Harvard University Press.

Kim, J. (2005). *Physicalism, or something near enough*. Princeton, NJ: Princeton University Press.

Kim, J. (2006). *The philosophy of mind*. Cambridge, MA: Westview Press.

Lewis, D. (1966). An argument from the identity theory. *Journal of Philosophy, 63,* 17–25.

Levine, J. (2003). *Purple haze*. Oxford, England: Oxford University Press.

Loar, B. (1981). *Mind and meaning*. Cambridge, England: Cambridge University Press.

Loar, B. (1997). Phenomenal states. In N. Block, O. Flanagan, & G. Guzeldere (Eds.), *The nature of consciousness* (pp. 597–616). Cambridge: MIT Press.

Lycan, W. (1996). *Consciousness and experience*. Cambridge, England: Cambridge University Press.

McGinn, C. (1989). Can we solve the mind-body problem? *Mind, 98,* 349–366.

McLaughlin, B. P. (1992). The rise and fall of British emergentism. In A. Beckerman, H. Flohr, & J. Kim (Eds.), *Emergence or reduction? Essays on the prospects of nonreductive physicalism* (pp. 49–93). Berlin, Germany: Walter de Gruyter.

McLaughlin, B. P. (2001). In defense of new wave materialism: A response to Horgan and Tienson. In B. Loewer & C. Gillett (Eds.), *Physicalism and its discontents* (pp. 317–328). Cambridge, England: Cambridge University Press.

McLaughlin, B. P. (2006). Type materialism for phenomenal consciousness. In S. Scneider (Ed.), *The Blackwell companion to consciousness* (pp. 431–444). Oxford, England: Basil Blackwell.

McLaughlin, B. P. (2010). Consciousness, type physicalism, and inference to the best explanation. *Philosophical Issues, 20,* 266–304.

McLaughlin, B. P. (2012). On justifying neurobiologicalism for consciousness. In S. Gozzano & C. Hill (Eds.), *New perspectives on type identity* (pp. 207–229). Cambridge, England: Cambridge University Press.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 4,* 435–450.

Nagel, T. (1986). *The view from nowhere.* Oxford, England: Oxford University Press.

Nagel, T. (1999). Conceiving the impossible and the mind-body problem. *Philosophy, 73,* 337–352.

Papineau, D. (2002). *Thinking about consciousness*. Oxford, England: Clarendon Press.

Plato. (1961). Phaedo. In E. Hamilton & H. Carins (Eds.), *Plato: The collected dialogues* (pp. 40–98). Princeton, NJ: Princeton University Press.

Putnan, H. (1968). Brains and behavior. In R. Butler (Ed.), *Analytical philosophy: Second series* (pp. 1–19). Oxford, England: Basil Blackwell.

Putnan, H. (1973). Psychological predicates. In W. H. Captain & Merill, D. D. (Eds.), *Art, mind, and religion* (pp. 37–48). Pittsburgh, PA: University of Pittsburgh Press.

Rosenberg, G. (2005). *A place for consciousness: Probing the deep structure of the natural world.* Oxford, England: Oxford University Press.

Russell, B. (1927). *The analysis of matter.* London, England: Kegan & Paul.

Ryle, G. (1949). *The concept of mind*. New York, NY: Barnes & Noble.

Seager, W. (2009). Panpsychism. In B. P. McLaughlin, A. Berckermann, & S. Walter (Eds.), *The Oxford handbook of philosophy of mind* (pp. 206–222). Oxford, England: Oxford University Press.

Searle, J. (1992). *The rediscovery of mind*. Cambridge, England: Cambridge University Press.

Skrbina, D. (2005). *Panpsychism in the West*. Cambridge: MIT Press.

Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review, LXVIII*, 141–156.

Strawson, G. (1994). *Mental reality.* Cambridge: MIT Press.

Strawson, G. (2006). *Consciousness and its place in nature: Does physicalism entail pansychism* (A. Freeman, Ed.). Exeter, England: Imprint Academics.

Strawson, G. (2008). *Real materialism and other essays*. Oxford, England: Oxford University Press.

Stubenberg, S. (1988). *Consciousness and qualia*. Amsterdam, Netherlands: John Benjamin.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, LIX,* 433–460.
Tye, M. (2000). *Consciousness, color, and content.* Cambridge: MIT Press.
Van Inwagen, P. (1975). Dualism and materialism: Athens and Jerusalem? *Faith and Philosophy, 4,* 475–488.

Mirror Neurons and Motor Cognition in Action Explanation

Mirror neurons are brain cells that seem to provide a functional matching between the motor properties of one's own actions and the sensory properties of the actions of other individuals. This entry presents the core neurological findings about *mirror neurons*, their distinctive properties, functioning, and varieties, and goes on to discuss various interpretations offered about the role of mirror neurons in motor cognition and action explanation.

Basic Findings

Basic findings about mirror neurons have been obtained during single-cell recordings. While there is a large imaging literature on mirror neurons in humans, this literature is inherently ambiguous because brain imaging measures vascular changes related to neural activity in a large number of neurons. In contrast, the evidence obtained with depth electrodes is quite compelling, and now such evidence is finally also available in humans.

Mirror neurons were originally discovered in an area of the monkey brain that controls hand and mouth movements. In this area (called F5 and located in the ventral premotor cortex), neurons fire when the monkey makes object-oriented actions, for instance, when it grasps, holds, and manipulates objects with the hand. Neurons in F5 also fire when the animal uses its mouth to bite food, to drink, or to make communicative facial expressions, such as lip smacking and when the monkey makes hand-to-mouth movements. Among the F5 motor neurons that fire when the monkey makes the actions just described, there are also some neurons that fire when the monkey is *not moving at all* but is *watching* another individual making the same or similar actions that achieve the same goal (e.g., grasping a piece of food with the hand or with the mouth). These neurons are called *mirror neurons* because it is as if the monkey is watching its own actions reflected by a mirror when watching another individual making actions.

Mirror neurons in area F5 are divided into two major classes: (1) *strictly congruent* mirror neurons (about one third of the recorded mirror neurons in F5), which fire for the same action, executed and observed, and (2) *broadly congruent* mirror neurons (about two thirds of the recorded mirror neurons in F5), which, on the other hand, fire for executed and observed actions that need not be identical but that achieve the same goal (e.g., grasping a peanut with the right hand or the left hand), or are somewhat "logically related," as, for instance, watching somebody placing food on the table and grasping the food. The fact that the majority of mirror neurons belong to the "broadly congruent" category suggests that—as a population—these cells can provide a flexible coding of the relationships between the actions of the self and the actions of others. So the mirror metaphor should not be taken too literally. These mirrors seem smart.

Mirror neurons in F5 have also the following properties: They do not fire at the sight of somebody else pantomiming an action; they fire, however, at the sound of an action (like breaking a peanut), even though the action is not seen at all; they also fire if the completion of the action cannot be seen (e.g., when a screen covers the completion of a grasping action; as long as the monkey knows that there is something to be grasped behind the screen, watching somebody reaching behind the screen will trigger the firing of the cell). Mirror neurons in F5 will also fire at the sight of somebody using a tool (e.g., pliers to grasp food), as long as the action has been repeatedly observed and it achieves a goal (like grasping food) that can be achieved with an action that the F5 neuron codes from a motor standpoint (like grasping).

Taken together, these properties suggest that what is mirrored by these cells is mostly the *goal* of the action. This is not surprising if one considers that even purely motor neurons in F5 seem to code mainly the goal of the action. Indeed, some purely motor neurons in F5 code for grasping an object with the left hand, the right hand, and even the mouth. Clearly, there are no commonalities between these three actions from the standpoint of body

part displacement or muscle contraction. However, all three actions share the same goal, grabbing the object. The most dramatic example supporting this concept comes from an experiment in which monkeys were trained to use pliers to grasp objects. Monkeys were trained to use normal pliers, which require closing the hand (i.e., flexing the fingers, as in a natural grasp) when grasping. Monkeys were also trained to use reverse pliers (which require the monkey to open the hand, by extending the fingers) to grasp objects. The same F5 cell would fire when the monkey used both normal and reverse pliers, even though the two actions required the contractions of completely opposite muscles of the hand, muscles that naturally operate in an agonist–antagonist fashion (when one contracts, the other relaxes, and vice versa).

Area F5 is anatomically connected with the anterior part of the inferior parietal lobule (called PF/PFG). Here, there are also neurons with properties virtually identical to F5 neurons, purely motor neurons and mirror neurons that mainly code for hand and mouth object-oriented actions. An experiment conducted in this brain area in monkeys demonstrated that the majority of neurons, both motor and mirror, that fire when the monkey grasps an object, code not for the grasping act per se but rather for the *intention* associated with the grasping action. Using simple contextual cues (the presence or absence of a container), the study tested two main intentions associated with the grasping action, *grasping to eat* and *grasping to place in a container*. The majority of the neurons recorded had differential discharges for the grasping action associated with the two different intentions, even though at the time of grasping the two intentions were obviously indistinguishable. This is evidence that a majority of these neurons code even *beyond the immediate goal* of the act (grasping that object). Perhaps the neuronal coding of these neurons should be conceived of as embedded in a chain of activation of different neurons that support the unfolding of coordinated actions that achieve specific intentions.

Recent studies have also shown that approximately 50% of mirror neurons will fire selectively depending on the proximity of the sector of space in which the observed action occurs. While phenomenologically we feel that the space surrounding us is "just one thing," the primate brain, including our own, contains at least two distinct maps of space. There is a map for peripersonal space, the space surrounding our own body, and a map for extrapersonal space, the space outside our reach. These space maps are "pragmatic maps" because they are dictated by the kind of actions that can be performed in the two different space sectors. Neurons supporting peripersonal space maps tend to code for reaching movements, for movements of the face and neck, and for coordinated actions defending the body. Neurons supporting extrapersonal space maps tend to code for eye movements. Some mirror neurons code for observed actions only when they occur in one of the two space sectors. Some mirror neurons code for observed actions happening in extrapersonal space and others for observed actions occurring in peripersonal space. This coding is also flexible and pragmatic. For instance, neurons coding for extrapersonal space maps do not fire for an action in peripersonal space. However, if a glass screen is interposed between the animal and the action occurring in peripersonal space, such that the animal can still see the action but cannot intervene on it, because of the glass screen (so that from a pragmatic standpoint it is as if the action occurs in extrapersonal space), then the neurons fire. This evidence clearly suggests that at least some of the coding that mirror neurons provide can be used for "online" social interactions.

Although for many years neurons with mirroring properties have been recorded only in the F5-PF/PFG network, recently a series of studies in monkeys and one case in humans have demonstrated mirroring responses in other cortical areas and for other types of movements. Three different labs have reported mirroring responses for reaching movements in the dorsal premotor and primary motor cortex. Furthermore, mirror neurons have been recorded in area LIP (lateral intraparietal cortex), an area that codes for the ocular movements that direct attention to novel objects in the environment. Mirroring in LIP may support joint attention, a fundamental building block in the development of social cognition.

Mirroring responses have also been observed in VIP (ventral intraparietal area), an area implementing peripersonal space maps and defensive body movements. VIP neurons respond to arm, face, and neck movements and tend to have bimodal receptive fields, responding to tactile stimulation and to the sight of three-dimensional (3-D) objects near the

body. The tactile and visual receptive fields are also spatially congruent. For instance, if a bimodal VIP neuron has a tactile receptive field on the forearm, it also responds to the sight of 3-D objects near the forearm, generally within the 20 to 30 centimeters of space surrounding the tactile receptive field. The strongest responses in these VIP neurons are elicited with objects moving toward the body, which suggests that VIP implements a map of the space surrounding the body that can be used to defend the body. Indeed, prolonged stimulations over cortical parietal sites around the intraparietal sulcus can elicit highly coordinated defensive movements. A recent study has demonstrated that some VIP neurons fire not only at the sight of a 3-D object near the tactile receptive field of the neuron, but also when the monkey observes a 3-D object nearing the corresponding body part (say the forearm) of another monkey. It is unclear whether the mirroring is here for the sensory event itself or for preparing a related defensive movement. The same neurons code both the sensory (sight, touch) and the motor (defensive action) event.

Auditory-vocal mirror neurons have been recorded in songbirds. These cells are active both when the songbird sings and when it is simply listening to the birdsong. They code specific sequences of the birdsong, and their activity is highly reproducible during singing and during listening. The motor aspect of these cells is demonstrated by the fact that their activity is unchanged while singing, even when the auditory feedback to the songbird is distorted. The existence of mirror neurons in songbirds demonstrates that neural mirroring is not a prerogative of primates and could be much more widespread in the animal kingdom than was initially thought.

Depth electrode recordings in presurgical patients have recently discovered mirror neurons in two human neural systems: the supplementary motor area (SMA) in the medial wall of the frontal lobe, and in two areas of the medial temporal lobe (MTL). To understand these findings, one must keep in mind that the placement of the electrodes in these studies is dictated only by clinical considerations, not research ones. Thus, recordings were obtained only from a small set of areas. None of these areas included human brain areas that are the homologs of the monkey brain areas in which mirror neurons were previously recorded. The SMA is a premotor region that is especially important for the selection and initiation of actions and for action sequences. However, the study that reported the human mirror neurons in SMA was not designed to investigate action sequences. Furthermore, the firing of the SMA mirror neurons coincided with the onset of the action, both when executed and when observed. Thus, it is reasonable to conclude that this population of mirror neurons enables mirroring for action onset.

The discovery of MTL mirror neurons in humans is surprising, because MTL is not a brain structure known for motor properties but rather for high-level visual properties and for memory properties. A potential interpretation of mirror neurons in MTL is that they mirror the *memory trace* of the observed action. That is, when I grasp a cup of coffee, my brain not only activates the motor plan for grasping the cup but also creates a memory of the grasping action. When I see somebody else grasping something, my brain not only reactivates the motor plans for grasping (using my premotor mirror neurons) but also reactivates the memory trace of my own grasping actions (using MTL mirror neurons).

Taken together, the data from single-cell recordings suggest a very rich mirroring of the actions of others. While the original findings pointed to mirroring mostly of the goals of grasping actions and some facial communicative gestures, the new data suggest that movement onset, the memory of the actions, reaching, defensive actions, and attentional eye movements, are also mirrored. These, obviously, are only those actions, or some of their specific aspects, that we know for sure are coded by neural mirroring. Given the variety of the empirical findings collected so far, it is not unlikely that all sorts of actions and their specific aspects are coded by mirror neurons.

Interpretation

While watching the actions of others, the observer activates—through mirror neurons—neural activity that is associated with her or his own actions. Thus, mirror neurons seem to enable an understanding "from within" of the actions of other people. Since actions are typically associated with intentions, feelings, and mental states, a corollary assumption of this interpretation is that mirror neurons make it possible to understand the mental states of others, thus enabling *empathy*.

This interpretation of mirror neuron activity maps well onto the assumptions of *motor cognition*, according to which cognitive processes are grounded

in our motor experience. Similarly, *embodied cognition* assumes that cognitive processes are grounded in our perceptual and motor experience. According to the ideomotor model of action, perception and action share common representational formats (mirror neurons are nice empirical evidence in its support). Thus, there are many similarities between the embodied and the motor cognition framework.

These theoretical frameworks challenge traditional models in classical cognitivism, according to which the operations of the mind are rule-based manipulations of symbols, largely detached from perceptual and motor experience. Inspired by these more traditional views of the mind, a series of objections have been raised against the "action understanding" hypothesis of mirror neurons. One objection is that people understand actions that they don't know how to perform. This objection, however, accepts a very limited notion of action understanding. Phenomenologically, motor experience enables a much richer understanding of perceived actions. As a tennis player, I am amazed at how little my family members understand of a tennis match we all watch in the living room. My richer understanding can't be enabled simply by more visual experience, since the visual experience on the court and in the living room are radically different. Mirror neurons would enable this deeper and richer form of action understanding.

Another objection is that the activation of mirror neurons during action observation is only epiphenomenal. There would be a higher-level understanding of observed actions that triggers—top down—a cascade of activations, including the activation of the motor plan to perform the observed action. This objection makes a specific prediction that is empirically testable: Visual activation should precede motor and mirror activation during action observation. Data in humans, however, have shown simultaneous co-activation of mirror and purely sensory neurons, rejecting this hypothesis.

More trivial objections are that neurological patients with motor deficits still understand the actions of others, that mirror neurons are simply "motor selection" neurons, and that they are blind to contextual cues that completely change the meaning of the observed action (consider Stevenson's fictional character Dr. Jekyll and his alter ego, the murderous Mr. Hyde, each intent on performing the same action with a scalpel on another person, but with completely different intentions).

The existence of some neurological patients with motor deficits who do understand actions is not problematic for the theory. Not all motor neurons are mirror neurons. However, the existence of some patients with motor deficits who also have action-understanding deficit is supportive of the action-understanding hypothesis of mirror neurons.

The hypothesis that mirror neurons are "motor selection" neurons is rejected by the fact that the majority of them (with the exclusion of the human SMA neurons described above) fire throughout the action, well after the action has been selected. The "blindness to contextual cues" hypothesis is rejected by the empirical study reported above on coding intentions using contextual cues.

A key theoretical issue that has been largely neglected until now is the role of control systems for mirroring in action understanding. Clearly, there are control systems for mirroring that prevent people from imitating whatever action they see. This is demonstrated by the imitative behavior of some neurological patients who after brain damage cannot help but imitate what other people do. It is unclear whether these control systems are general-purpose cognitive control mechanisms or control mechanisms dedicated to mirror neurons. Also unclear is the role of these control systems in action understanding. Perhaps they may be critical for the understanding of alternative actions, such as, for instance, fleeing when attacked.

Marco Iacoboni

See also Embodied Cognition; Empathy; Joint Attention and Social Cognition; Simulation Theory; Social Cognition; Social Neuroscience; Social Perception

Further Readings

Caggiano, V., Fogassi, L., Rizzolatti, G., Thier, P., & Casile, A. (2009). Mirror neurons differentially encode the peripersonal and extrapersonal space of monkeys. *Science, 324,* 403–406. doi:10.1126/science.1166818

Cisek, P., & Kalaska, J. F. (2004). Neural correlates of mental rehearsal in dorsal premotor cortex. *Nature, 431,* 993–996. doi:10.1038/nature03005

Dushanova, J., & Donoghue, J. (2010). Neurons in primary motor cortex engaged during action observation. *European Journal of Neuroscience, 31*(2), 386–398. doi:10.1111/j.1460–9568.2009.07067.x

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science, 308*(5722), 662–667. doi:10.1126/science.1106138

Gallese, V., Gernsbacher, M. A., Heyes, C., Hickok, G., & Iacoboni, M. (2011). Mirror neuron forum. *Perspectives on Psychological Science, 6*(4), 369–407. doi:10.1177/1745691611413392

Iacoboni, M. (2008). *Mirroring people*. New York, NY: Farrar, Straus & Giroux.

Ishida, H., Nakajima, K., Inase, M., & Murata, A. (2010). Shared mapping of own and others' bodies in visuotactile bimodal area of monkey parietal cortex. *Journal of Cognitive Neuroscience, 22*(1), 83–96. doi:10.1162/jocn.2009.21185

Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: A critique. *Trends in Cognitive Science, 9*(1), 21–25. doi:10.1016/j.tics.2004.11.003

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology, 20,* 750–756. doi:10.1016/j.cub.2010.02.045

Prather, J. F., Peters, S., Nowicki, S., & Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature, 451*(7176), 305–310. doi:10.1038/nature06492

Shepherd, S. V., Klein, J. T., Deaner, R. O., & Platt, M. L. (2009). Mirroring of attention by neurons in macaque parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 106*(23), 9489–9494. doi:10.1073/pnas.0900419106

Tkach, D., Reimer, J., & Hatsopoulos, N. G. (2007). Congruent activity during action and action observation in motor cortex. *Journal of Neuroscience, 27*(48), 13241–13250. doi:10.1523/JNEUROSCI.2895–07.2007

Modal Logic and Intentional Agency

Actions such as raising one's arm, switching on a computer, or killing someone are investigated in several areas of philosophy, among others in the philosophy of action, the philosophy of mind, and the philosophy of law. Through the analogy between actions and software programs, the concept of agency is also relevant in computer science, in particular in artificial intelligence (AI), multi-agent systems, and theoretical computer science. In the second half of the 20th century, philosophers and logicians like G. H. von Wright and Stig Kanger started to use "modal logic" as a tool for modeling intentional action and related concepts such as *omission*, *ability*, and *opportunity*.

The Two Approaches

There are two general views of the logical form of action sentences: (1) the "action as result" view and (2) the "action as means + result" view. The entry reviews these two approaches and recent developments.

According to the "action as result" view, an action can be identified with a state of affairs that is brought about by an agent. In other words, the action of an agent can be identified with the result that the agent causes. Examples of states of affairs are that an arm is up, that a computer is turned on, that somebody is dead, and so on. So my action of switching on the computer is identified with my bringing about the state of affairs that the computer is started, and my action of killing someone is identified with my bringing about the state of affairs that someone is dead. Examples of modal logics of action taking this perspective are the Logic of Seeing-To-It That (STIT), proposed by Nuel Belnap, John Horty and colleagues, and the Logic of Bringing-It-About-That (BIAT), proposed by Kanger and recently elaborated by Dag Elgesem. For instance, STIT logic has modal operators [*i*: **stit**], where *i* is an agent name. The STIT formula [*i*: **stit**]*p* means that "the agent *i* sees to it that the proposition *p* is true."

The alternative view, "action as means + result," focuses on both the result and the means by which an action is performed. It is about sentences such as "I bring it about that the computer is on by toggling the switch" or "I bring it about that someone is dead by poisoning him." Examples of modal logics of action taking this perspective are variants of Propositional Dynamic Logic (PDL), where an agent argument is added to events (which were introduced in theoretical computer science), as well as Krister Segerberg's Logic of "Bringing-About-By-Doing." For example, there exist variants of PDL that have modal operators of the form [*i*: α], where *i* is an agent name and α is an action name. The formula [*i:* α]*p* means that "after the agent *i* performs the action α, it is the case that *p*."

Recent Developments

In the past two decades, there have been several proposals to extend the modal logic analysis of action sentences expressing intentional action in both the "action as result" tradition and the "action as means + result" tradition.

Logicians in the "action as result" tradition, such as Jan Broersen (a contributor to this encyclopedia), Caroline Semmling, and Heinrich Wansing, have proposed to distiguish the fact that an agent *accidentally* sees to it that a given state of affairs is true from the fact that an agent *intentionally* sees to it that a given state of affairs is true. A way to capture this distinction is by extending STIT logic with special operators for intentional action of the form [*i*: **istit**], where [*i:* **istit**]*p* means that "the agent *i* intentionally sees to it that the proposition *p* is true." A fundamental relationship between these operators and standard STIT operators is expressed by the following logical axiom:

$$[i\text{: }\mathbf{istit}]p \rightarrow [i\text{: }\mathbf{stit}]p,$$

which means that if an agent intentionally causes *p* to be true, then he causes *p*. One can easily distinguish intentional action captured by the operator [*i*: **istit**] from nonintentional (accidental) action captured by the formula

$$[i\text{: }\mathbf{stit}]p \wedge \neg[i\text{: }\mathbf{istit}]p.$$

That is, "accidentally causing *p*" just means "causing *p* without intentionally causing *p*." Another way to capture the distinction between intentional action and accidental action is by extending STIT logic with modal operators for goal. These operators are called *volitional* modal operators (expressing the idea of intending or willing to act), in opposition to epistemic and doxastic modal operators, which are studied in the context of epistemic logic (the modal logic of knowledge) and doxastic logic (the modal logic of belief), respectively. The generic form of these goal operators is $Goal_i$, where $Goal_i\ p$ means that "the agent *i* wants (or wishes) *p* to be true." Intentional action is then captured by the formula

$$[i\text{: }\mathbf{stit}]p \wedge \mathrm{Goal}_i\ [i\text{: }\mathbf{stit}]p,$$

whereas nonintentional (accidental) action is captured by the formula

$$[i\text{: }\mathbf{stit}]p \wedge Goal_i\ [i\text{: }\mathbf{stit}]p.$$

That is, "intentionally causing *p*" means "causing *p* with the goal of causing *p*," whereas "accidentally causing *p*" means "causing *p* without the goal of causing *p*."

Logicians in the "action as means + result" tradition have been more interested in studying the concept of *intention*. The seminal work in this area is Philip R. Cohen and Hector J. Levesque's article "Intention Is Choice With Commitment." Starting from Michael Bratman's philosophical theory of action, Cohen and Levesque have analyzed intention as a version of PDL in which time is assumed to be linear. On their account, intentions are defined in several steps in relation to the concept (and corresponding modal operator) of "goal"; that is, intention is viewed as a strongly realistic preference: Among the worlds that are (epistemically) possible for an agent, there is a subset that the agent prefers. According to Cohen and Levesque's analysis, a given agent *i* has the intention that *p* if and only if the agent *i* has the persistent goal that *p* and believes he can achieve that goal by an action α of his, where a persistent goal that *p* is a goal that is kept by the agent until it is either fullfilled or believed to be out of reach.

In the article "A Logic of Intention and Attempt," Emiliano Lorini and Andreas Herzig complemented Cohen and Levesque's approach by integrating the concept of an attempt to perform an action. The concept of "attempt" is indeed crucial in order to relate the agent's intention (the mental world) with the performing of the action (the physical world). The central principles here are that "if an agent *i* has the intention to perform a given action α of his, then he attempts (or tries) to perform the action α" and that "if an agent *i* tries to perform a given action α of his and has the capability of doing α, then he will successfully perform the action α."

Emiliano Lorini

See also Action, Philosophical Theory of; Agency; Deontic Logic and Agency; Intentionality

Further Readings

Belnap, N., Perloff, M., & Xu, M. (2001). *Facing the future: Agents and choices in our indeterminist world*. New York, NY: Oxford University Press.

Bratman, M. (1987). *Intentions, plans, and practical reason*. Cambridge, MA: Harvard University Press.

Broersen, J. (2011). Making a start with the STIT logic analysis of intentional action. *Journal of Philosophical Logic, 40,* 399–420.
Cohen, P. R., & Levesque, H. J. (1990). Intention is choice with commitment. *Artificial Intelligence, 42,* 213–261.
Elgesem, D. (1993). *Action theory and modal logic* (Unpublished doctoral dissertation). Norway, Oslo: University of Oslo.
Fitch, F. B. (1963). A logical analysis of some value concepts. *Journal of Symbolic Logic, 28,* 135–142.
Harel, D., Kozen, D., & Tiuryn, J. (2000). *Dynamic logic*. Cambridge: MIT Press.
Horty, J. F., & Belnap, N. (1995). The deliberative STIT: A study of action, omission, and obligation. *Journal of Philosophical Logic, 24*(6), 583–644.
Kanger, S. (1971). New foundations for ethical theory. In R. Hilpinen (Ed.), *Deontic logic: Introductory and systematic readings* (pp. 36–58). Dordrecht, Netherlands: D. Reidel.
Lorini, E., & Herzig, A. (2008). A logic of intention and attempt. *Synthese, 163*(1), 45–77.
Meyer, J. J. C., van der Hoek, W., & Van Linder, B. (1999). A logical approach to the dynamics of commitments. *Artificial Intelligence, 113*(1–2), 1–40.
Segerberg, K. (1989). Bringing it about. *Journal of Philosophical Logic, 18,* 327–347.
Segerberg, K. (1992). Getting started: Beginnings in the logic of action. *Studia Logica, 51*(3–4), 347–378.
Semmling, C., & Wansing, H. (2008). From BDI and STIT to BDI-STIT logic. *Logic and Logical Philosophy, 17,* 185–207.
von Wright, G. H. (1963). *Norm and action*. London, England: Routledge & Kegan Paul.

Models in Science

The *model-oriented stance* in the philosophy of science places models at center stage in our analysis of scientific practice and focuses on issues to do with their multivarious nature(s) and role(s). Many of the relevant examples in this entry have been drawn from the physical sciences, but there are obvious ways in which the discussion can be extended to the social sciences.

The Nature of Models

Models come in all shapes and sizes, as it were. Some are physically constituted, as in Francis Crick and James Watson's famous "tinplate and wire" model of DNA; others are theoretical, as in the Lotka-Volterra model of predator–prey relationships. Within the latter category, some may be quantitative and others qualitative, and combinations of both may feature in scientific investigations. What distinguishes a model from a theory, broadly construed, may not always be clear (and indeed, there are many examples of scientists using these terms interchangeably). However, typically, models incorporate certain simplifications and idealizations that render them more tractable, either computationally or otherwise. Despite the idealized aspects, models *represent* the system concerned. To do so in the relevant respects, and to the requisite degree, the model must be *similar* to the system, in at least some regard.

This focus on the nature of models as representations provides a general framework that is capable of embracing all the very different kinds of models one finds in the sciences. Here we are talking about the nature of models as representational devices at the level of scientific practice. One can then shift up a level and consider ways of representing models (and theories) themselves, at the level of the philosophy of science. At this level, the variety of different kinds of models has often been cited as blocking the construction of any such unitary framework. However, one can regard all such kinds—whether physical, theoretical, or whatever—as instantiating certain kinds of structures, and this structural or model-theoretic approach, then, offers a formal account of representation and also broadens the manner in which models represent, from similarity to the sharing of various kinds of structural features.

Furthermore, this last emphasis on the structural commonalities between models and the systems being modeled helps articulate the various roles that models play with regard to understanding, explanation, and drawing inferences about the system.

The Role of Models

By emphasizing only certain features of the system of interest, a model may then allow certain inferences to be more easily drawn. By drawing such inferences within the model and then transferring these inferences to the system—via the structural commonalities between the model and the system—we perform what is known as *surrogative reasoning*. This ability to support such inferences illustrates the representational power of models, but of course, it would not be possible to make these inferences unless the

appropriate relationship of structural similarity held between the model and the system in the first place.

It is, in part, because of this ability that models have been taken to possess a form of functional "autonomy" in the sense of acting as the locus for knowledge claims, providing the basis for further developments, and so on. A glance at the relevant literature in any particular domain of science will reveal models acting as "autonomous agents" in this uncontentious sense. The stronger claim that models are autonomous by virtue of being constructed independently of theory is more problematic and must be judged on a case-by-case basis.

Furthermore, it is by virtue of such functional autonomy that models act as *mediators* between theory and phenomena. Thus, for a given theory to explain and be confirmed by the relevant phenomena, certain simplifications and idealizations may typically have to be introduced. This mediating function may be extended to accommodate the sense in which models may help relate different kinds of explanation within a field. So with regard to the different levels at which explanations in the social sciences are articulated, assumptions associated with certain models at one level may intrude into another, allowing for interesting interlevel relationships.

Again, shifting to the level of the philosophy of science and of social science itself, this intrusion of models from one level to another, and interlevel relationships in general, can be usefully and formally represented by the model-theoretic approach. The adoption of such a formal framework can then underpin claims about the nature of such interrelationships, both between and within levels as well as between theories, models, and the relevant phenomena.

It is also by virtue of the structural relationships that hold between models and the relevant systems that certain kinds of models can be used in computer simulations, both allowing scientists to explore certain dynamical features of systems that might be difficult or impossible to access otherwise and also suggesting revisions to the model or even alternative models entirely. Given the complex nature of many systems studied in the social sciences, model-based computer simulations offer an attractive option.

And, of course, it is this complexity that raises problems when it comes to capturing the regularities that scientists are interested in via the kinds of laws that one finds in the physical sciences. Here, models may offer a useful substitute, and in the philosophy of biology, for example, a broadly model-theoretic stance is now common. The extent to which such a stance can be extended into the social sciences hinges on a number of methodological issues but remains an open and interesting question.

Steven French

See also Agent-Based Modeling and Simulation in the Social Sciences; Explanation, Theories of; Game-Theoretic Modeling; Idealization in Social Scientific Theories; Mathematical Models, Use in the Social Sciences; Models in Social Science; Multi-Agent Modeling; Scientific Method

Further Readings

Brante, T. (2010). Perspectival realism, representational models, and the social sciences. *Philosophy of the Social Sciences, 40,* 107–118.

Contessa, G. (2011). Scientific models and representation. In J. Saatsi & S. French (Eds.), *The continuum companion to the philosophy of science* (pp. 120–137). New York, NY: Continuum.

da Costa, N. C. A., & French, S. (2003). *Science and partial truth: A unitary approach to models and scientific reasoning*. Oxford, England: Oxford University Press.

French, S. (2010). Keeping quiet on the ontology of models. *Synthese, 172*(2), 231–249.

Morgan, M. S., & Margaret Morrison, M. (Eds.). (1999). *Models as mediators: Perspectives on natural and social science* (pp. 38–65). Cambridge, England: Cambridge University Press.

Odenbaugh, J. (2008). Models. In S. Sarkar & A. Plutynski (Eds.), *Blackwell companion to the philosophy of biology* (pp. 506–524). Oxford, England: Blackwell.

Suárez, M. (2010). Scientific representation. *Philosophy Compass, 5,* 91–101.

Suppes, P. (1960). A comparison of the meaning and uses of models in mathematics and the empirical sciences. *Synthèse, 12,* 287–301.

Models in Social Science

The social sciences use models in order to understand phenomena, to predict trends (e.g., in mortality, morbidity and fertility), and to inform policy actions. For instance, a social scientist may build a model to explain why variations in maternal education are consistently associated with variations in child mortality in developing countries. She may build a model to predict how lung cancer rates

will change if tobacco consumption is considerably reduced. Or she may build a model that supports a policy action on unemployment benefits in order to reduce unemployment rates. Those are empirical models, namely, models that employ data. There are also models that do not employ data and that are instead more theoretical or formal in character. These are more often used in economics and are intended to reconstruct economic processes and behaviors. An example is Thomas Schelling's model of segregation or Milton Friedman's hypothesis of permanent income.

This entry focuses on empirical models used in the social sciences, analyzes their central features, presents central epistemological issues related to them, and shows the relation between models and theories in natural science as well as in social science.

Empirical models can be quantitative or qualitative, depending on the methods used to analyze data. Quantitative models make prominent use of probability and statistics in order to analyze large amounts of data, typically numerically collected; a consequence of this approach is that the larger the size of the sample, the more reliable the analysis. Qualitative models, instead, employ methods for the collection and analysis of data that are independent of statistics and probability. They typically focus on the narrative provided by the chosen interviewees and pay particular attention to their language, background, beliefs, and other contextual aspects.

Another important distinction to be made is the following. In quantitative analysis, models can be associational or causal, depending on their goal. While associational models only aim to provide a faithful description of a phenomenon, causal models go further in searching for the causes of that phenomenon. For instance, an associational model may describe how obesity rates vary according to socioeconomic classes. A causal model, to follow up on the same example, would go further in stating whether and why socioeconomic differences *cause* obesity.

Models are evaluated according to their *validity* or *invalidity*. The locus classicus is the seminal 1979 book by Thomas D. Cook and Donald T. Campbell. There is a vivid debate both in social science methodology and in philosophy about the methods for ensuring the validity of models. There are two main types of validity: *internal* and *external*. Internal validity concerns whether a relationship between two variables is causal, or whether from the absence of a relationship between the two variables we can infer absence of causality. External validity concerns the possibility of generalizing a presumed causal relationship across different times, settings, or populations.

There are also philosophical issues making the notion of validity problematic. For instance, scientists disagree whether internal or external validity is more important; whether external validity is needed or achievable at all in social science; whether validity concerns the results, the model, or the data; and so on. Validity is a notion that certainly needs further investigation and conceptual clarification. Despite all the difficulties that undermine any attempt to provide an account of validity, there is at least one reason to keep the notion. The validity *of a model* allows us to endorse an overarching view where establishing whether *X* causes *Y* depends on the evaluation of the *many* elements of a model: background knowledge, assumptions, data, statistical tests, and so on. This is at variance with the approach typically endorsed in philosophical analyses: saying what makes a causal claim (i.e., *X* causes *Y*) true. Granted, in ordinary language the issue is usually to establish whether the claims we made are true or false based on ordinary causal assessment. For instance, competent speakers would think that it is true that I wouldn't have missed the train had I heard the alarm clock ringing this morning. However, in science, establishing causal relations goes far beyond finding the "facts" or "objects" that make a causal claim true. It instead involves evaluating the whole modeling procedure from beginning to end. This implies that we can challenge the validity of a model at different stages: poor theoretical support, bad data collection, failure of statistical tests, and so forth.

Much debate in the philosophical literature is devoted to the distinction between *models* and *theories*. In social contexts, the distinction has a clearer borderline. Models are used to analyze data and participate in developing general theories. For instance, models on migration that are used to analyze empirical data in different countries and at different times can contribute toward developing a general theory of migratory movements. However, the achievement of general and robust theories in the social sciences is much more difficult than in, say, physics, because human behavior is highly mutable across time, space, and culture. This implies that findings in social

science are highly context dependent. Yet it does not mean that they are not objective or reliable. Typically, (empirical) modeling of phenomena in social science participates in building theories. Economics is perhaps an exception. According to classical economics, economic theory dictates the mechanism (also called the data-generating process) underlying the data, and modeling is just about empirical testing.

Federica Russo

See also Agent-Based Modeling and Simulation in the Social Sciences; Causal Explanation, in Philosophy of Science; Causation, Philosophical Views of; Causation in the Social Sciences; Chicago School (Economics); Mathematical Models, Use in the Social Sciences; Models in Science

Further Readings

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.

Franck, R. (Ed.). (2002). *The explanatory power of models*. Dordrecht, Netherlands: Kluwer.

Morgan, M. S., & Morrison, M. (1999). *Models as mediators*. Cambridge, England: Cambridge University Press.

Russo, F. (2009). *Causality and causal modelling in the social sciences: Measuring variations*. New York, NY: Springer.

Wunsch, G. (1988). *Causal theory and causal modelling*. Leuven, Belgium: Leuven University Press.

Modernity

This entry charts the conceptual and historical unfolding of the notion of modernity, central to the development of both philosophy and the social sciences.

It is customary to locate the origins of *philosophical* modernity in Europe in the period from the middle of the 17th to the end of the 18th century. From René Descartes to Immanuel Kant, scholars posited the individual human being as the knowing subject who, through the combination of freedom and reason, develops an understanding of the world that is radically different from all the earlier ones. In parallel, a particular philosophy of *political modernity* emerged that held that the rules for a peaceful life in common would best—or only—be established through a contract between free and reason-endowed individuals (from Thomas Hobbes's *Leviathan* to the combination of individual freedom and collective self-determination in John Locke). Even though they co-emerged, the foundational philosophy and the political philosophy of modernity proved to stand in tension with each other, which explains much of the relation between such philosophies and the social sciences that were to emerge in the course of the 19th century.

In both its strands, such thinking provided indeed a new and radical expression of the human condition. It would be erroneous, however, to equate it with the understanding of modernity that became dominant in—and characteristic of—Europe and the West, to see it as the philosophy of European modernity *tout court*. Rather, the positing of the autonomous, reason-endowed individual became an extreme point in a tension-rich field of interpretations of modernity.

From the late 18th century onward, the aforementioned modern self-understanding was criticized with regard to its concept of the isolated individual as the knowing and acting subject under conditions of modernity. In response, concepts of intersubjectivity and of situated freedom were proposed to replace the grounding of modern action in the individual. The key contribution came from G. W. F. Hegel at the beginning of the 19th century, and it keeps being referred to as central by contemporary philosophers of modernity such as Jürgen Habermas, Axel Honneth, Robert Pippin, and Charles Taylor.

After the French Revolution, more specifically, political thinkers came increasingly to hold that the founding of the modern polity in the free individual and her capacity for reason alone would be insufficient. The bond of reason and interest, it was argued, was too thin and the extension of social relations in emerging commercial and industrial society too wide to sustain such a polity. In response to this political problem, the recourse to intersubjectivity, however, seemed less convincing than with regard to a basic social ontology. The idea that the social world was created and maintained through interactions between situated persons, and not the least through bonds of communication that would create an action-guiding "public opinion," lost plausibility for large societies like postrevolutionary France and most of her European neighbors.

In turn, a new thinking about social bonds that existed regardless of the will of the singular human beings and had the potential of integrating large collectivities and/or introducing a dynamics of change was proposed, and this in two basic versions. Some authors, maybe most prominently Gottfried Herder, held that cultural-linguistic bonds existed between human beings in the form of shared values and beliefs and commonality of language. Modern polities would be sustainable if they were created by and with speakers of the same language. This thought fed into the national liberation and unification movements and, later, the nationalism of 19th- and 20th-century Europe. Other authors assumed that the rise of commercial and industrial society would extend relations of social interest across large groups, either through mutual dependence in a division of social labor or through the commonality of class position. The latter position was central to Karl Marx's thought, the former to Émile Durkheim's.

This reconceptualization of social bonds stood at the origins both of the social sciences and of the social transformations that went in parallel from the later 19th century onward. Collective concepts such as nation, class, and society became key concepts of the social sciences and referred to social phenomena that were created to deal with the insufficiencies of liberal and instrumental thinking, which as such formed the core of economic theorizing. The social sciences witnessed their disputes on method with foci on the relation between holist (collectivist) and atomist (individualist) ontologies, on the one hand, and abstract reasoning and situated (historicist) reasoning, on the other. Social life witnessed a transformation of restricted liberal modernity into an organized modernity that was based on collective conventions and institutions of various kinds framed by exactly such concepts as nation, class, and society.

In the early 20th century, in particular after World War I, the tensions within these intellectual and political constellations proved unbearable. Politically, the critique of the consequences of political and economic liberalism led to collective existentialisms, of which Nazism, based on a concept of nation, and Stalinism, based on a concept of class, were the most pronounced examples. Philosophically, the critique of modernity and its intellectual tool sets culminated in calls for radical renewal, such as in Martin Heidegger's work or in the early Frankfurt school of Theodor W. Adorno and Max Horkheimer.

After World War II, a less tension-ridden and more inclusive version of liberal modernity, devoid of philosophy and flanked by technocratic social sciences, seemed possible. Our contemporary thinking, however, is marked by a new round of critique, often retrieving older resources, that emerged during the 1960s and culminated toward the end of the 1970s. In their broadest meaning, the terms *poststructuralism* and *postmodernism* stand for a philosophical critique of the "discourse of the human sciences" (Michel Foucault and Jacques Derrida) of the preceding two centuries and for a sociopolitical critique of organized modernity and its claims of having made functional and normative accomplishments unavailable to other societies elsewhere and in the past (Jean-François Lyotard).

These debates and—maybe more importantly—recent transformations of the world that have seen the rise of non-Western capitalisms and the radical critique of Western modernity arising in various forms out of the situation of postcoloniality have reopened the debate about modernity. Most significantly, they have reintroduced a discussion about the ambiguous nature of "occidental rationalism," of which Max Weber had identified the claim of having both local origins and universal significance, thus inviting comparison between the presumed Western and other trajectories of modernity. This has entailed, conceptually, the recognition of a wider interpretative field of philosophy of modernity than has long been assumed and of a plurality of forms of political modernity that may coexist in the current global context.

Peter Wagner

See also Enlightenment, Critique of; Intersubjectivity; Multiculturalism; Philosophes, The; Philosophy of Politics, History of; Postcolonial Studies; Postmodernism; Social Contract Theories

Further Readings

Derrida, J. (1978). *Writing and difference*. London, England: Routledge.

Habermas, J. (1987). *The philosophiocal discourse of modernity*. Cambridge: MIT Press.

Taylor, C. (1995). *Philosophical arguments*. Cambridge, MA: Harvard University Press.

Wagner, P. (2001). *Theorizing modernity*. London, England: Sage.

Modularity of the Mind

The concept of mental modularity plays a prominent role in both philosophy and psychology. As originally articulated by Jerry Fodor in the early 1980s, it characterizes a type of information-processing mechanism that is likely to figure in relatively low-level parts of the functional architecture of the mind. More recent developments of modularity theory, beginning in the early 1990s, loosened this characterization, thereby fostering the idea of the mind as more pervasively modular in organization.

This entry briefly surveys these two phases of thinking about modularity and points to their significance for the philosophy of social science.

Fodorian Modularity

According to the conception of modularity articulated by Jerry Fodor, a module is a component of the functional architecture of the mind that exhibits a cluster of interrelated features. The hallmarks of Fodorian modularity are domain specificity, mandatory operation, limited central accessibility, fast processing, informational encapsulation, "shallow" outputs, fixed neural architecture, characteristic and specific breakdown patterns, and characteristic ontogenetic pace and sequencing. Since each of the features on this list admits of degree, modularity does so as well, but a system will count as modular if it exhibits enough of them to a significant degree. And while each feature is characteristic of modular systems, some features are more central than others, and one of them, in particular—informational encapsulation—is the most central of all.

A system is "informationally encapsulated" to the extent that, in the course of processing its inputs, it has restricted access to information stored outside the system itself. For example, the visual system has limited access to what the perceiver explicitly believes about the character of a visual stimulus, since such information is not stored in a proprietary database. This is suggested by the persistence of visual illusions, such as the Müller-Lyer illusion, in which a pair of equal parallel lines appear to differ in length even after the viewer is convinced of the contrary (see Figure 1).

This particular phenomenon also illustrates the related concept of "cognitive impenetrability."

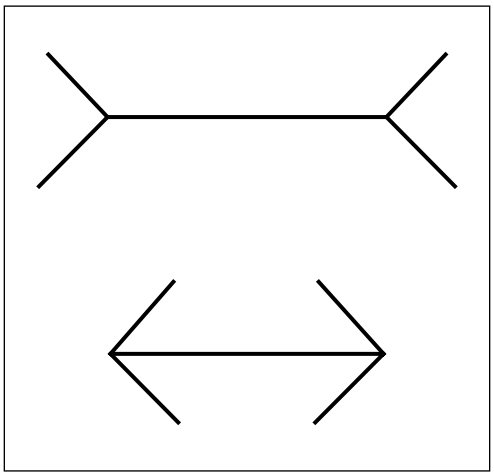


Figure 1 The Müller-Lyer Illusion

Source: Author.

Cognitive impenetrability is a species of encapsulation, namely, encapsulation relative to what the perceiver believes about the world (as opposed to what the deliverances of the visual system might lead her to believe).

Informational encapsulation lies at the heart of modularity, largely because of its explanatory priority vis-à-vis several other features on Fodor's roster. Consider, for instance, the fact that modular processes are typically fast (a half-second or less from start to finish) and mandatory (not under conscious control) and their outputs are "shallow" (computationally cheap and informationally sparse). These features can be seen, at least to some extent, as by-products of a module's restricted informational purview during processing. The same goes for other hallmarks of modularity, such as neural localizability ("hardwiring"), functional dissociability, and innateness.

Is the Mind Really Modular?

With this conception of modularity in play, a natural question to ask is this: How much of the mind is modular in this sense? Fodor's answer is clear: not very much, at least insofar as one thinks of higher cognitive functions, such as decision making, problem solving, and planning, as paradigmatic of the mental. On Fodor's view, only the "sensing and acting" parts of the mind—that is, the relatively low-level systems dedicated to perception and motor control—exhibit much in the way of modularity; the "thinking" parts of the mind do not. This is

due largely to the fact that systems responsible for higher cognition, unlike sensorimotor systems, are *not* informationally encapsulated. Indeed, it seems almost constitutive of higher cognitive operations that they draw freely on widely disparate information stores in the brain. Hence, to the extent that modular systems are explanatorily tractable in a way that nonmodular systems are not, the prospects for a successful science of the mind appear to be somewhat limited. (This *pro tanto* claim is roughly equivalent to what Fodor calls the "First Law of the Non-Existence of Cognitive Science.")

These considerations lead naturally to a second question: Is there some other sense, less restrictive than Fodor's, in which the mind might be said to be modular? The answer from psychology, and *evolutionary psychology* in particular, is a resounding *yes*. On this view, modularity is no longer a peripheral feature of our mental architecture but a pervasive one.

Massive Modularity

The suggestion that the mind might be "massively modular," that is, modular through and through, has relatively little going for it so long as one thinks of modularity along Fodorian lines. The principal reason for this, as noted above, is Fodor's insistence that modularity requires informational encapsulation. Once we relax this requirement, the class of cognitive systems that might count as modular expands considerably.

Accordingly, proponents of massive modularity like Peter Carruthers emphasize other facets of Fodorian modularity, such as domain specificity, automaticity, and inaccessibility of a system's operations to first-person report. Of these features, *domain specificity* tends to get the spotlight. Like encapsulation, domain specificity is a measure of a system's informational reach, but in a different (and logically independent) sense. A system is domain specific insofar as the range of inputs that can turn it on, or initialize its operations, is relatively narrow. For example, the systems responsible for recognizing faces and voices, seeing colors, and analyzing shapes all count as domain specific, since each of these systems is effectively attuned to a particular class of stimulus. That sort of selective attunement, however, does not require informational encapsulation, or conversely.

Much of the enthusiasm for massive modularity stems from the application of *evolutionary* ideas to cognitive psychology. Massive modularists like Leda Cosmides and John Tooby, for example, have argued for the view as follows. First, the human mind is a product of natural selection. Second, selective fitness requires the cognitive ability to solve a range of adaptive problems, such as finding food and shelter, selecting mates, and negotiating social exchanges. Third, adaptive problems can be solved more quickly, efficiently, and reliably by modular cognitive systems than by nonmodular ones. Hence, a modular architecture for cognition (i.e., not just perception and motor control) is more plausible on evolutionary grounds than its nonmodular counterpart.

Lessons for the Philosophy of Social Sciences

The significance of massive modularity theory, and of modularity theories more generally, for the philosophy of social science is difficult to overestimate. For present purposes, two points will have to suffice.

First, given the long-standing debate within social science regarding the relative contributions of nature and nurture to human development (or, as Steven Pinker puts it, the relative plausibility of nativism vis-à-vis empiricism, or the "Blank Slate" model), the affinity between modularity and innateness makes modularity a natural focus of investigation within the philosophy of social science.

Second, insofar as philosophers of social science are concerned with developing an account of psychological explanation—and to the extent that psychological explanation is construed as a species of mechanistic explanation—the distinction between modular and nonmodular mechanisms will likely figure in any such account. This is reflected in recent debates about the functional architecture of social-cognitive capacities such as reasoning about social exchanges (e.g., cheater detection) and reading other minds.

Philip Robbins

See also Classical Computationalism, Connectionism, and Computational Neuroscience; Cognitive Sciences; Evolutionary Psychology; Sociobiology

Further Readings

Carruthers, P. (2006). *The architecture of the mind*. Oxford, England: Oxford University Press.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J.

Tooby (Eds.), *The adapted mind* (pp. 163–228). Oxford, England: Oxford University Press.
Fodor, J. A. (1983). *The modularity of mind*. Cambridge: MIT Press.
Pinker, S. (2002). *The blank slate: The modern denial of human nature*. New York, NY: Viking.
Robbins, P. (2007). Minimalism and modularity. In G. Preyer & G. Peter (Eds.), *Context-sensitivity and semantic minimalism* (pp. 303–319). Oxford, England: Oxford University Press.

Money

This entry offers a historical and conceptual overview of social scientific and philosophical accounts of money.

Money has long intrigued philosophers and social scientists. Scholars have debated the nature of money at least since Aristotle, who argued in *The Nicomachean Ethics* that money is not natural but a creature of law or convention whose value is subject to alteration. This distinction between the intrinsic value of the monetary medium and the social and legal context on which its definition as money depends continues to characterize present-day debates about the difference between substance- and token-money. It is also important to arguments about the origins of money, central banking, metallism and chartalism, the gold standard, and the role of money in relation to international trade. There are two main schools of thought in monetary theory. On the one side are "chartalist" or "claim" theories of money, which follow Max Weber, Georg F. Knapp, and John Maynard Keynes in arguing that money's value depends on political and legal authority. On the other side are commodity theories, which argue that money's value resides either in its own intrinsic properties as a medium or in its relationship with an underlying commodity, such as gold. In addition, there are "cultural" accounts of money, which suggest that both claim and commodity theories of money underestimate the importance of money's more symbolic properties, which are rooted in the social contexts in which it is used.

In economics, money is defined as a medium of exchange, a store of value, and a unit of account. But the topic demands deeper questioning, throwing up puzzles about the nature of time, number, and space, and for this reason, philosophers, historians, anthropologists, sociologists, political scientists, and geographers have all made major contributions to our understanding of the nature and workings of money. The richest works of monetary scholarship therefore straddle several fields. Perhaps most notably, Georg Simmel's *Philosophy of Money* was a work of sociology as well as philosophy that started out in the early 1890s as an article on the psychology of money. In the book itself, Simmel insisted that "not a single line" of his 600-page book was about economics; indeed, he took a quite different view of the value of money. By his reckoning, money represents an abstract idea of value that is underwritten by "society," depending on a form of trust he likened to religious faith.

Simmel can be counted among a number of modern thinkers—including Karl Marx—who feared that money's increasing use in society would have a corrosive impact on social relationships, emptying them of deeper significance and meaning, reducing qualitative connections to quantitative ones, and rendering our relations to each other increasingly functional and utilitarian. Sociologists and anthropologists who argue that money richly conveys meanings and symbolism that are derived from the social and cultural context of its use have since challenged this view. While anthropologists have debated whether some of this richness is lost in modern economic systems, in sociology the work of Viviana Zelizer is notable for arguing that whenever we use money—whatever the context—we "earmark" it and, thereby, render it distinctive and personal. For anthropologists and sociologists working in this tradition, money is therefore not the abstract and impersonal medium of calculation that classical economic and social thinkers held it to be.

Empirically, scholars have been debating how to define "new" and "alternative" monetary media, such as LETS tokens, Time Dollars, commercial reward schemes, and e-money, alongside the "hybrid" monetary systems that operate in marginal regions of the global economy, such as Melanesia and Africa. Broadly speaking, two countervailing themes stand out in the academic literature. On the one hand, some scholars are focusing on the fact that large-scale currencies like the U.S. dollar are circulating more and more outside the borders of their issuing states and, in some cases, are actually replacing smaller currencies. This process constitutes a trend

toward the increasing *homogeneity* of money. From the perspective of political economy, key drivers of monetary homogenization are state currency competition (dollarization) and regional factors (monetary union). From a Marxist perspective, economic factors, such as the balance of global capital flows, are crucial and account for what the geographer David Harvey calls the "spatio-temporal fix." On the other hand, there are scholars who argue that the range of monetary forms in circulation that are not state-issued currency is expanding, primarily outside the mainstream banking system. This, they claim, constitutes a trend toward increasing the *diversity* of money. The drivers of monetary diversification are connected on one side to financialization and, on the other, to local money and social lending.

The diversification of monetary instruments within the finance industry has been driven by increasingly sophisticated risk management, the efforts of the banks to evade national regulation regimes, and the intense pace of technological innovation in the construction of financial instruments—an increasing tendency to profit from money itself. As for local currencies and social lending networks, monetary diversification is largely driven by economic necessity (financial exclusion), economic advantage (transaction costs), and political commitments to community-related and ecological goals (including more radical programs informed by neoanarchist beliefs). Both trends suggest a decline in the influence of states over the world's money flows; indeed, there is sometimes an observable connection between monetary homogenization and diversification: The widespread use of local currency schemes during the Argentine economic crisis of 1999 to 2002 was an expression of both trends.

The view that states are losing their monopoly over the production and management of money has been reinforced by the development of financial instruments such as derivatives, which some scholars believe demonstrate all the essential features of money. Today's monetary flows connect nodal points such as global cities—with their major financial centers and business districts—along with other major rallying points for money, such as the international art markets. In effect, these centers serve as bottlenecks in the global circulation of money. The major consequence of these flows, according to some geographers, is money's deterritorialization: We are witnessing the end of money's geography. This analysis is central to Michael Hardt and Antonio Negri's *Empire*, which draws on the work of the French philosophers Gilles Deleuze and Félix Guattari to examine the "rhizomatic mobility" of money-capital, whose control is no longer delimited by national borders or by traditional international boundaries. However, the recent financial crisis, which saw governments taking part-ownership of troubled banks, suggests that the argument that states have become irrelevant to the world of global finance is premature. The recent emergence of sovereign wealth funds as a major force in the world's monetary and financial system complicates this picture still further.

Nigel Dodd

See also Economic Anthropology; Markets and Economic Theory; Philosophy of Economics, History of; Simmel's Philosophy of Society

Further Readings

Bryan, D., & Rafferty, M. (2006). *Capitalism with derivatives: A political economy of financial derivatives, capital and class*. London, England: Palgrave-Macmillan.

Cohen, B. J. (2004). *The future of money*. Princeton, NJ: Princeton University Press.

Corbridge, S., Martin, R., & Thrift, N. (Eds.). (1994). *Money, power and space*. Oxford, England: Blackwell.

Davies, G. (2002). *A history of money*: *From ancient times to the present day*. Cardiff, Wales: University of Wales Press.

Deleuze, G., & Guattari, F. (2004). *Capitalism and schizophrenia: Vol. 1. Anti-Oedipus*. London, England: Continuum Books. (Original work published 1972 in French as *Capitalisme et schizophrénie: Vol. 1. L'Anti-Œdipe*, Paris, France: Les editions de Minuit)

Dodd, N. (1994). *The sociology of money*. Cambridge, England: Polity.

Dodd, N. (2011). "Strange money": Risk, finance and socialized debt. *British Journal of Sociology, 62*(1), 1–20.

Eagleton, C., & Williams, J. (2007). *Money: A history*. London, England: British Museum Press.

Gilbert, E., & Helleiner, E. (Eds.). (1999). *Nation-states and money: The past, present, and future of national currencies*. London, England: Routledge.

Hardt, M., & Negri, A. (2000). *Empire*. Harvard, MA: Harvard University Press.

Hart, K. (2001). *Money in an unequal world*. New York, NY: Texere.

Harvey, D. (2007). *The limits of capital*. London, England: Verso.
Ingham, G. (2004). *The nature of money*. Cambridge, England: Polity.
Simmel, G. (2004). *The philosophy of money*. London, England: Routledge.
Zelizer, V. (1994). *The social meaning of money*. New York, NY: Basic Books.

Montesquieu and the Rise of Social Science

No theorist of the French Enlightenment contributed more to the development of the social sciences than Charles Louis de Secondat, Baron de Montesquieu (1689–1755). Although the phrase *la science social* was only first used near the end of the 18th century and initially had public policy connotations, it is clear that Montesquieu anticipated developments we now associate with the scientific investigation of human societies.

This entry gives an account of Montesquieu's contribution to the rise of the social sciences qua "sciences" as we have come to know them, as well as of the particular methodological precepts and social scientific axioms he himself formulated to explain both specific government types or the general spirit of nations and the rationality of human social and political action.

In suggesting a linkage between regime types, expanse of territory, and the underlying climatological, geographical, and cultural factors, he demarcated the field of political sociology. In connecting religious beliefs and practices to the underlying social conditions and psychological needs, he anticipated the emergence of the sociology and psychology of religion. In comparing European and Asian practices, he contributed to the founding of comparative law and comparative religion. By focusing attention on a wide range of customs displayed by preliterate cultures, he contributed to the birth of ethnography and anthropology. Moreover, Montesquieu made important contributions to history, political science, sociology, and economics by writing a causal account of the fall of Rome; revising the regime typology inherited from Plato and Aristotle to take account of *how*, rather than *by whom*, power is exercised; constructing ideal types for aiding analysis of governmental form; and analyzing the social and political effects of commerce as well as the diverse sources of the wealth of nations.

Montesquieu understood that *social facts* occupy a realm apart from governmental, psychological, biological, or chemical facts, and he therefore sensed that society presents its own particular subject matter for the social scientist to explore. Moreover, he assumed that certain supposed attributes of human nature cannot explain the enormous variety of customs, manners, and morals that can be observed in various cultures around the world. In addition, like the French sociologist Émile Durkheim substantially later, Montesquieu understood that societies are natural, *organic* entities that by no means originate in an act of social contracting, as so many other political theorists had claimed. Montesquieu also perceived that it is the presence of internalized *values* and *beliefs*—"virtue" for republics, "honor" for monarchies, and "fear" for despotisms—that explains the durability of the social structures underlying governmental forms. Society is *not* the product of individual minds isolated from one another. Rather, society results from *interaction* between individuals and is greatly influenced not only by the principle of government but also by what he termed the "general spirit" (*esprit général*) that, once formed, imposes itself on individual minds and shapes human behavior.

Practitioners of social science assume the *rational* explicability of social phenomena. There can be no science if there are no patterns to discover and if no laws can be devised that describe and predict social behavior. Montesquieu was convinced that rather than acting capriciously, human beings often make rational choices influenced by underlying causes. Thus, he asserted in the Preface to *The Spirit of Laws* (1989), "I began by examining men, and I believed that, amid the infinite diversity of laws, and mores, they were not led by their fancies alone" (p. xliii). Following this train of thought, he remarked in Book I of his treatise that "blind fate" cannot be the cause of "all the effects that we see in the world." He acknowledged that the "intelligent world" displays less regularity than the natural world since human beings possess free will and can err through ignorance, but he believed that the patient observer can nonetheless discern the causes that render human behavior explicable. Thus, he remarked in his notebooks, *Mes Pensées*, "When a law seems bizarre, and

there is no sign that the Legislator has had an interest in making it so . . . , one ought to believe that it is more reasonable than it appears, and that it is based on a sufficient reason" (Montesquieu, 1991, p. 589).

Inspired by the successes of Kepler, Galileo, and Newton in formulating the laws governing matter and the solar system, Montesquieu assumed that similar progress could be made in discovering the constant relations (*rapports*), or *natural regularities*, rendering human social behavior explicable. He concluded that when subjected to a given set of circumstances, human beings will react in predictable ways, and this means that *societal behavior can be studied by the methods we now associate with social science*.

The accuracy of Montesquieu's assertions is less important than their source, namely, his conviction that political, social, and economic practices can be traced to discernible *causes*. It was this viewpoint that encouraged him to formulate a number of now famous social science axioms. He concluded, for example, that despotism arises in countries characterized by extreme heat and vast territorial expanse, where subjects lack energy to resist and where only fear of harsh measures can render subjects on the periphery loyal to a distant central government. Democratic republics will spring up, on the other hand, in extremely small, frugal societies that nourish equality and where direct citizen involvement in lawmaking enables self-sacrificing political virtue to take root. Slavery will be found in hot climates, where labor is burdensome, while polygamy will exist where gender imbalances make women plentiful and men scarce. Even religious phenomena, Montesquieu believed, can be traced to specific, causal influences. Protestantism, for example, flourished in northern Europe because the cold climate there gives the people a spirit of independence and liberty, whereas the people of southern Europe were better suited to Catholicism, which has a more authoritarian structure.

Montesquieu's conviction that social, political, economic, and religious phenomena are explicable in terms of underlying *causes* is apparent in the very title of his posthumously published *Essay on Causes Affecting Minds and Characters*. In this essay, composed between 1736 and 1743, he analyzed the physical causes (*causes physiques*) of human behavior before proceeding to an analysis of the moral, social, and intellectual influences (*causes morales*) shaping societal practices. His list of physical causes included climate, soil, topography, and terrain, while his enumeration of moral causes included maxims of government, laws, education, religion, modes of subsistence, and the customs and manners that shape behavior in a given society.

Foreshadowing Émile Durkheim's later emphasis on the "collective conscience" of a society, Montesquieu believed that every nation develops a "general spirit," or overall disposition, which is the result of all the major influences shaping national character, including climate, religion, laws, maxims of government, precedents, morals, and customs. In sorting out the relative strength of these influences Montesquieu developed what sociologists term a *functionalist* perspective, envisioning social phenomena as part of a larger social totality. The stronger the influence of one particular factor in a nation's overall disposition, the weaker the others are likely to be. For example, when the moral precepts of a religion are strict, the civil and criminal laws will generally be less severe. Montesquieu understood that nations are unique blends of idiosyncratic influences, and he concluded that the general spirits of nations represent a subtle blending of multiple causes that gives each people a distinctive national character. Every nation develops its own "natural genius," which it is best not to alter. Even faults, Montesquieu suggested, should normally be left unaddressed since they are part of an overall blend of influences producing the general spirit, or the common culture to which a nation has grown accustomed.

Montesquieu was by no means a strictly empirical social scientist refraining from *moral judgments*. He combined the descriptive, scientific outlook that prompted August Comte, Émile Durkheim, and Raymond Aron to consider him the originator of sociology with a *normative*, more traditional natural-law perspective that led him to forthrightly condemn evils such as despotism, slavery, torture, religious fanaticism, ruthless colonial practices, cruel and excessive punishments, and legal systems requiring self-incrimination. Much more than the other theorists of his day, however, he set out to discover the reasons why societies develop particular laws, customs, religions, and institutions. Hence, he remarked in his Preface to *The Spirit of Laws*, "I did

not draw my principles from my prejudices but from the nature of things" (p. xliii).

Montesquieu's sociological approach and appreciation of cultural diversity contributed to his belief that there is no ideal form of government that every people should adopt. The right government for a given people, he remarked in Book I of *The Spirit of Laws*,

> is the one whose particular arrangement best relates to the disposition of the people for whom it is established. . . . Laws should be so appropriate to the people for whom they are made that it is very unlikely that the laws for one nation will suit another. (p. 8)

Laws should reflect the particular circumstances present in a society, including "the nature and principles of the government"; "the physical aspect of the country"; the "climate," "terrain," and "extent of the country"; the socioeconomic stage a people has attained; "the degree of liberty that the constitution can sustain"; religious beliefs; wealth, or lack thereof; population size; commerce; mores; and manners (p. 9).

Given Montesquieu's interest not so much in specific laws and legal systems but rather in the underlying influences on laws, which constitute the "spirit of the laws," it is not surprising that Durkheim accorded him a key role in defining the subject matter of social science and that Aron regarded him as not just a precursor of sociology but as the veritable founder of that discipline that seeks to discover the structure underlying human societies. Montesquieu was clearly seeking what Durkheim would later refer to as "social facts" and "the real relations among things." Moreover, like Durkheim, Montesquieu did not perceive society as merely the individual reified. In addition, he grasped that individuals are socialized by the forces and influences that society brings to bear on them. Thus, in the Preface to *The Spirit of Laws*, he referred to "man" as "that flexible being who adapts himself in society to the thoughts and impressions of others" (pp. xliv–xlv). And, finally, through his construction of republics, monarchies, and despotisms as ideal types, he anticipated Max Weber's extensive use of ideal types as aids to sociological understanding.

It is certainly not surprising that François Quesnay and other physiocrats in France as well as numerous other theorists of the Enlightenment—including Jean d'Alembert, A. R. J. Turgot, and the Marquis de Condorcet in France; David Hume, Adam Ferguson, Lord Kames, John Millar, and Adam Smith in Scotland; and J. G. von Herder in Germany—drew much inspiration from Montesquieu for their own scientific investigations of social phenomena. Although Montesquieu lived at a time when the various disciplines of social science had not yet been delineated, he was clearly working on many of the problems that still engage those who seek to explore the behavior of human beings in society in a scientific manner.

David W. Carrithers

See also Causation in the Social Sciences; Durkheim's Philosophy of Social Science; Holism, in the Social Sciences; Individualism, Methodological; Law, Social Phenomenon of; Philosophes, The; Philosophy of Politics, History of; Philosophy of Sociology, History of; Rationality and Social Explanation; Structural Functionalism, in Social Theory

Further Readings

Aron, R. (1965). *Main currents in sociological thought* (Vol. 1). New York, NY: Basic Books.

Carrithers, D. (1998). The enlightenment science of society. In C. Fox, R. Porter, & R. Wokler (Eds.), *Inventing human science: Eighteenth-century domains* (pp. 232–270). Berkeley: University of California Press.

Carrithers, D. W., Mosher, M. A., & Rahe, P. A. (Eds.). (2001). *Montesquieu's science of politics: Essays on* The Spirit of Laws. Lanham, MD: Rowman & Littlefield.

Durkheim, É. (1960). *Montesquieu and Rousseau: Forerunners of sociology*. Ann Arbor: University of Michigan. (Original work published 1893)

Durkheim, É. (1973). Sociology in France in the nineteenth century. In R. Bellah (Ed.), *Emile Durkheim on morality and society* (pp. 3–22). Chicago, IL: University of Chicago Press. (Original work published 1900)

Durkheim, É. (1982). *Rules of sociological method*. New York, NY: Free Press.

Montesquieu. (1989). *The spirit of laws* (A. M. Cohler, B. C. Miller, & H. S. Stone, Eds.). New York, NY: Cambridge University Press.

Montesquieu. (1991). *Pensées: Le Spicilège* [Thoughts: Scrapbook] (L. Desgraves, Ed.). Paris, France: Robert Laffont. (Original work published 1934)

Stark, W. (1965). *Montesquieu: Pioneer of the sociology of knowledge*. Toronto, Ontario, Canada: University of Toronto.

Moral Cognitivism

Moral cognitivism is a family of views within the branch of ethical theory known as *metaethics*. Metaethics studies questions about the fundamental nature and status of moral beliefs, judgments, and theories. Rather than taking a stand, say, on the conditions under which actions are morally required or permitted, or states of affairs morally good or bad, metaethical views instead focus on whether such substantive ethical claims can be true, whether and how they can be known, whether they are in some way rationally compelling, and whether and how they motivate us.

The defining features of any version of moral cognitivism are (a) its commitment to the *truth-evaluability* of moral propositions and assertions and (b) its view of *moral assertions* as intended primarily to accurately represent the moral truth. The first commitment is straightforward: All moral cognitivists think that moral claims can be assessed as either true or false. But this is not enough to define one as a cognitivist, since a number of sophisticated noncognitivists nowadays share this view.

Where cognitivists and their counterparts disagree is with regard to what is happening in central cases of moral judgment. Cognitivists claim that moral agents are ordinarily attempting in their moral judgments to make factual assertions that seek to depict a world in which things exemplify a variety of moral features. On the cognitivist view, moral judgments are attempted reports of the moral facts. Of course there are nonstandard cases of moral utterances, in which agents are lying or insincerely parroting the views of others. But the cognitivist says that most moral judgments, and all central cases of moral judgment, amount to *assertions* of the putative truth about what is right, good, admirable, and so on.

This view is opposed to *noncognitivism*, which holds that moral judgments are centrally judgments in which a person expresses some nonrepresentational attitude toward some nonmoral state of affairs. The core idea in noncognitivism is something like this: *There are no values in the world*, really. There is, roughly, the world that science tells us about, and our emotional or affective responses to that world. Our moral judgments enshrine those responses and use moral vocabulary to convey our sentiments. Moral judgments are not really in the business of describing the way the world is. They are rather in the business of expressing our emotional responses to a value-free world.

Moral cognitivists believe that people are attempting to speak the truth when they issue their moral judgments. But some cognitivists deny that there is any truth to be had in morality. These cognitivists are known as *error theorists*. According to error theorists, try as we might to get things right in our moral claims, we invariably fail, since nothing is morally right or wrong, good or bad. The basic idea behind error theories is that there is some fundamental assumption underlying all moral views—different error theorists may disagree about just which assumption this is—and that this assumption is false. People try to state the truth when making moral judgments. And they always fail, hence the error.

Error theorists are the pessimists within the moral cognitivist family. By far the greater number of cognitivists have instead endorsed the idea that some moral claims are true and that we can know that they are true. We can cognize about—that is, think and reason about—morality and sometimes come to correct moral answers in the correct way, backed by the correct reasons and arguments. This optimistic assessment is shared by all moral cognitivists except error theorists.

If we leave error theories aside, all other forms of moral cognitivism can be classed as versions of either *moral constructivism* or *moral realism*. Constructivist theories are defined by their commitment to the view that fundamental moral truths are in some way constructed from the attitudes or responses of a duly specified agent or set of agents. Realist theories deny this constructive role—realists are united in their view that fundamental moral truths are not made true by virtue of having been ratified, endorsed, or implied by the attitudes or responses of any given individual or set of individuals.

Constructivist theories range from various forms of relativism all the way to ideal-observer theories. Starting on the more "subjective" end of the spectrum, different versions of relativism insist that what counts as the correct moral theory—the correct theory that specifies the necessary and sufficient conditions for an action's being morally right or good—may properly differ from person to person, or from society to society. At the more "objective"

end, ideal-observer theories will claim that the fundamental moral truths are those that reflect the choices or attitudes of agents who are far smarter and more rational than any actual agent. Different versions of ideal-observer theories can be developed, each distinguishing itself from the others on the basis of which characteristics (and conditions of choice or response) are part of the idealization.

Realist theories are primarily divided between *nonnaturalist* and *naturalist* versions. Naturalist moral realists think of ethics as continuous with the natural sciences and think that a posteriori, empirical investigation is the proper way to discern all the substantive moral truths there are. Nonnaturalists reject this picture and resist the assimilation of moral inquiry to scientific inquiry. Nonnaturalists regard moral properties as sui generis, fundamentally unlike those that are the subject matter of the natural sciences. Being morally required, for instance, is essentially a normative feature of actions, and its normativity distinguishes it from the sorts of properties—such as being round or being red—that are the proper subject matter of the sciences. Or so nonnaturalists will argue.

In sum, all moral cognitivist theories share a commitment to the view that people are standardly trying to speak the *truth* when issuing moral judgments. Error theorists claim that everyone fails in such efforts, since, as error theorists see things, no moral claims are true. Constructivists and moral realists reject the attribution of fundamental error to morality, and each offers his or her own distinctive take on the status of the ultimate standards of morality.

Russ Shafer-Landau

See also Neuroethics; Realism and Anti-Realism in the Social Sciences; Reflective Equilibrium; Relativisms and Their Ontologies; Truth, Philosophical Theories of

Further Readings

Korsgaard, C. (1996). *The sources of normativity*. Cambridge, England: Cambridge University Press.

Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. New York, NY: Penguin.

Railton, P. (1986). Moral realism. *Philosophical Review, 95,* 163–207.

Shafer-Landau, R. (2003). *Moral realism: A defence*. Oxford, England: Oxford University Press.

Multi-Agent Modeling

This entry explains the importance and application of multi-agent modeling in social explanation and contrasts it with the kindred but distinct field of *cognitive architectures*. It also draws a parallel with the traditional sociological approaches, like those of Max Weber or Alfred Schütz, which can be further illuminated by these recent methods of computerized social simulation. Multi-agent modeling and social simulation have become an increasingly important research methodology for the social sciences.

Computerized Social Simulation of Agency

The notions of "agent" and "agency" have had a major role in framing research in the social and behavioral sciences. Influenced by computer science (including distributed artificial intelligence, computer networking, etc.), multi-agent modeling or *social simulation*, that is, simulation of social processes and phenomena on the basis of models of multiple autonomous individual agents, has become a significant aspect of the social sciences.

In this context, agents are *computational entities* each of which presumably represents an individual person. From their interactions, complex patterns may emerge, leading to various representations of social phenomena. Thus, the interactions among multiple computational agents provide potential explanations for the corresponding social phenomena.

Multi-Agent Social Simulation

Multi-agent social simulation has seen tremendous growth in recent decades. Researchers hoping to go beyond the limitations of traditional approaches to the social sciences have increasingly turned to agent-based social simulation for studying a wide range of theoretical and practical issues.

Traditionally, two approaches dominate the social sciences. One approach centers on the construction of mathematical models of social phenomena, usually expressed as a set of closed-form mathematical equations. Such models may be mathematically elegant but with limited expressive power. Deduction may be used to find the consequences of assumptions. The other approach may be more *qualitative* and *conceptual*, with which insights are obtained

by generalizations from observations in an informal (nonmathematical) way.

The new approach of multi-agent modeling (social simulation) has emerged relatively recently. It involves computational modeling and simulation of social phenomena based on multiple interacting agents. It starts with a set of detailed assumptions (in the form of rules, mechanisms, or processes) regarding agents and their interactions. Simulations lead to data that can be analyzed to come up with useful generalizations. Thus, simulations are useful as an aid to developing theories, or even as theories themselves.

One of the first uses of multi-agent models in the social sciences was by Robert Axelrod, in his study of evolution of cooperation. In this early work, computational simulations were used to study strategic behavior in the iterated Prisoner's Dilemma game. Even today, this work is still influencing research in various fields. In the mid-1980s, artificial-life modeling emerged, the idea of which was to simulate life to understand the basic principles of life. This led to the application in social simulation of many interesting ideas, such as complexity, evolution, self-organization, and emergence. These ideas influenced social scientists in developing and conducting social simulations. Recently, another topic area appeared dealing with the formation and the dynamics of social networks (i.e., social structures connected through social familiarities ranging from casual acquaintance to close familial bonds).

Cognitive Architectures

Related to social simulation but separately, psychologically realistic computational models of agents have been developed in cognitive science. In cognitive science, they are often known as *cognitive architectures*, that is, the essential structures and processes of cognition/psychology in a broadly scoped, domain-generic computational cognitive model. There are reasons to believe that computational cognitive architectures can play a significant role in multi-agent social simulation. Social simulation can benefit from incorporating cognitive architectures, because they provide a realistic basis for modeling individual agents.

For the field of multi-agent social simulation, the use of cognitive architectures leads to an interesting kind of explanation of social phenomena—*cognitively based explanation* of social phenomena. Social processes ultimately rest on the actions and decisions of individuals, and thus understanding the mechanisms of individual cognition/psychology can potentially lead to better theories describing the aggregates of multiple individuals. In traditional, early treatments of social explanation qua "understanding," the central desideratum was to try and understand the social agents' own conception of "meaningful behavior" and, more generally, the contextual meaning. Getting to know cognitive structures of agents' mental make-up was thus already on the agenda—even if in a different terminology. Max Weber, for example, pointed out that being different from the physical sciences, the social sciences need to gain an "empathetic understanding" of the "inner states" of social actors (their psychology/cognition), thus gaining an understanding at both the level of causation and the level of "meaning" (i.e., cognition/motivation). Alfred Schütz, for another example, attempted to understand the construction of social reality from the point of view of an individual, in terms of meaningful actions, motivations, and social relationships. Thus, a realistic agent model, incorporating realistic tendencies, inclinations, and capabilities of individual agents on the basis of contemporary cognitive sciences, can serve as a solid basis for understanding the interaction of individuals and social processes. What this boils down to is cognitive social simulation (as termed and argued for by Ron Sun), as opposed to mere agent-based social simulation. Cognitive architectures may be an important centerpiece of this enterprise.

So far, however, the two fields of multi-agent social simulation and cognitive architectures have developed rather separately from each other, with some notable exceptions. Much of the work in social simulation, for the sake of ease of programming and running simulation, assumed rudimentary cognition on the part of the agents.

Issues addressed thus far by multi-agent modeling have been diverse. They include, for example, social beliefs, norms, language evolution, resource allocation, traffic patterns, social cooperation, tribal customs, culture formation, stock market dynamics, group interaction and dynamics, organizational decision making, organization design, and countless others.

In all, multi-agent modeling and simulation have become an increasingly important research

methodology for the social sciences. They are now widely used for testing theoretical models (e.g., for investigating their properties when analytical solutions are not possible). A simulation may serve as a theory (or an explanation of a social phenomenon) by itself.

Ron Sun

See also Agency; Agent-Based Modeling and Simulation in the Social Sciences; Artificial Intelligence; Cognitive Sciences; Game-Theoretic Modeling; Mathematical Models, Use in the Social Sciences; Models in Social Science

Further Readings

Axelrod, R. (1984). *The evolution of cooperation*. New York, NY: Basic Books.

Gilbert, N., & Doran, J. (1994). *Simulating societies: The computer simulation of social phenomena*. London, England: University College.

Kohler, T., & Gumerman, G. (Eds.). (2000). *Dynamics in human and primate societies*. New York, NY: Oxford University Press.

Moss, S., & Davidsson, P. (Eds.). (2001). *Multi-agent-based simulation*. Berlin, Germany: Springer.

Sun, R. (2002). *Duality of the mind*. Mahwah, NJ: Lawrence Erlbaum.

Sun, R. (Ed.). (2006). *Cognition and multi-agent interaction: From cognitive modeling to social simulation*. New York, NY: Cambridge University Press.

MULTICULTURALISM

Multiculturalism has a number of meanings within and beyond social science. For some it refers to the sociological fact that different ethno-cultural groups are present in a single society. For others it refers to a society where individuals are presented with a smorgasbord of cultural choices and pick and mix as they choose, while no distinct groups are dominant. One should particularly beware of texts that define multiculturalism without reference to any of the work of its advocates, as is now common in public and even academic discourse, as this usually means that a caricature is being presented—typically, "cultural separatism" and "cultural relativism," neither of which has ever or rarely been argued for by multiculturalists.

Thus, multiculturalism is a site where two principal epistemological issues in the social sciences take on tangible forms: On the one hand, the individualism/holism divide is tested in concrete ways, while, on the other, the multifarious ways in which multiculturalism is defined reenact the essential contestability and definitional multiplicity of social and political concepts.

This entry focuses on one of the core meanings: The political accommodation of cultural minorities, a movement in the latter part of the 20th century, which filled some of the space that accommodation of the working classes had previously occupied for a century or more. Even within this restricted scope, in both theoretical and policy discourses, however, multiculturalism has different meanings in different places—and this is crucial for the social sciences. For example, in North America, its referents include groups with territorial claims, such as Native Peoples and the Quebecois, even though these groups want to be treated as "nations" within a multinational state, rather than merely as ethnocultural groups in a mononational state. In Europe, groups with such claims, like the Catalán and the Welsh, are thought of as nations, and multiculturalism has a more circumscribed meaning, referring to a postimmigration urban mélange and the politics it gives rise to. While in North America, language-based ethnicity is seen as the major political challenge; in Western Europe, the conjoining of the terms *immigration* and *culture* usually invokes the large, newly settled Muslim populations. Sometimes, usually in America, political terms such as multiculturalism and "rainbow coalition" are meant to include all groups marked by "difference" and historic exclusion, such as women and gays.

This coalitional meaning derives from the fact that the ethnic assertiveness associated with multiculturalism has been part of a wider political current of "identity politics" that emerged in the 1960s and transformed the idea of equality as sameness to equality as difference—or, in a related conceptualization, adding the concept of respect or "recognition" of a group identity to the older concept of equality as the equal dignity of individuals. Black power, feminist, and gay pride movements challenged the ideal of equality as assimilation and contended that true equality was when stigmatized groups could define themselves in positive terms, thus challenging an individualist ideal of color-blind, culture-neutral politics with the critique that ethnicity and culture

cannot be confined to some so-called private sphere but shape political and opportunity structures in all societies. It is the theoretical basis for the conclusion that allegedly "neutral" liberal democracies are part of a hegemonic culture that systematically de-ethnicizes or marginalizes minorities—hence the claim that minority cultures, norms, and symbols have as much right as their hegemonic counterparts to state provision and to be in the public space, to be recognized as groups and not just as culturally neutered individuals.

The prominence of political theory in multiculturalism is also partly to be understood in terms of the internal dynamic within the discipline. John Rawls's *Theory of Justice*, published in 1971, is the founding text in the modern revival of normative Anglo-American political theory. It promised a philosophically grounded, systematic answer to questions of distributive justice in societies such as the contemporary United States, which were assumed to be characterized by value pluralism. Subsequent debate, including Rawls's reformulation of his own position, focused not on his conclusions about distribution but on his assumptions about rationality and value pluralism. The generation of political theorists following Rawls thus have come to define their questions more in terms of the nature of community and minority rights than in terms of distributive justice, no less than their social-theory peers defined it in terms of difference and identity rather than class conflict, and in each case the intellectual framework lent itself to multiculturalism, even when the term itself was not favored. While for most political theorists academic liberalism has been the primary reference point, Bhikhu Parekh has offered a philosophical multiculturalism grounded in an analysis of human nature and culture and elaborating the intrinsic value of diversity as more fundamental than the accommodation of minorities.

One of the most fundamental divisions among scholars concerns the nature of "cultural groups." In sociocultural studies, groups are seen as having internal differences, hierarchies, gender inequality, and dissent; and culture is always fluid and subject to varied influences, mixtures, and change. To think otherwise is to "essentialize" groups such as Blacks, Muslims, Hispanics, and so on. Political theorists, on the other hand, continue to think of cultural groups as socio-political actors that may bear rights and have needs that need to be institutionally accommodated. This approach challenges the view of culture as radically unstable and primarily expressive by putting moral communities at the center of a definition of "culture." Empirical studies, however, suggest that both these views have some substance. For while many young people, from minority and majority backgrounds, do not wish to be defined by a singular ethnicity but wish to actively mix and share several heritages, there is simultaneously development of distinct communities, usually ethno-religious and sometimes seeking corporate representation.

Since the events of 9/11 and their aftermath, it is Muslims who have become the focus of discourse about minorities in the West. This is partly an issue of security, but more generally, it is accompanied by a "multiculturalism is dead" rhetoric and is found mostly in Western Europe. Muslims, it is argued, are disloyal to European states, preferring segregation and sociocultural separatism to integration; they are illiberal on a range of issues, most notably the personal freedom of women and homosexuality; and they are challenging the secular character of European political culture by thrusting religious identities and communalism into the public space. The prohibiting of Muslim identity in public space has so far been taken furthest in France, where, in 2011, Parliament banned the wearing of the face veil in public places.

Thus, if multiculturalism was initially about indigenous peoples, Black pride, and bilingualism, it is currently challenging ideas about the place of religion in public life, and so an interrogation—sometimes a reassertion—of a secularism that most western Europeans thought was settled and hegemonic. What unites the earlier and the current phase of multiculturalism is the aspiration to create new forms of internally plural national identities that nurture a sense of belonging among minorities as well as historic majorities. While the appeal of multiculturalism as a public policy has suffered considerable political damage, the intellectual and policy argument that multiculturalism is a valuable means of remaking public identities in order to achieve an equality of citizenship that is neither merely individualistic nor premised on assimilation remains powerful and is unlikely to be erased. It is in this way that multiculturalism appears, again, to be a privileged

site within which the individualism/holism divide at the center of the social sciences takes various yet concrete forms.

Tariq Modood

See also Cultural Studies; Holism, in the Social Sciences; Individualism, Methodological; Norbert Elias: Process of Civilization and Theory of Sciences; Oppression; Postcolonial Studies; Postmodernism; Power

Further Readings

Kymlicka, W. (1995). *Multicultural citizenship: A liberal theory of minority rights*. Oxford, England: Oxford University Press.

Modood, T. (2007). *Multiculturalism: A civic idea*. Cambridge, England: Polity Press.

Parekh, B. (2000). *Rethinking multiculturalism: Cultural diversity and political theory*. Cambridge, MA: Harvard University Press.

Taylor, C. (1994). The politics of recognition. In A. Gutmann (Ed.), *Multiculturalism: examining the politics of recognition* (pp. 25–74). Princeton, NJ: Princeton University Press.

Young, I. M. (1990). *Justice and the politics of difference*. Princeton, NJ: Princeton University Press.

Mutual Beliefs

The notion of mutual belief (also called *common belief*) has been extensively studied in the past four decades. It plays a central role in several areas of research, ranging from social philosophy and linguistics to game theory, theoretical computer science, and distributed artificial intelligence. This entry gives an account of the essential (formal and substantive) features of mutual belief and the crucial role it plays in analyses of joint action, group activity, cooperation, conversation, and other such social contexts in which *collective* items (collective agency, collective reasoning, collective intentionality, etc.) are the defining features.

Formally, *mutual belief* is distinguised from *shared belief* in the following way. We say that the agents in a set of agents G *share the belief* that a certain fact p is true if and only if each of the agents in G individually believes that p is true. We then say that the agents in G have the *mutual belief* that p is true if and only if the agents in G share the belief that p is true for every order k. That is, the agents in G share the belief that p is true, the agents in G share the belief that the agents in G share the belief that p is true, and so on ad infinitum. This is also called *iterative definition* of mutual belief and has been distinguished from the *fixed-point definition*, which can be stated as follows: The agents in G have the *mutual belief* that p if and only if every agent in G believes that p and every agent in G believes that the agents in G have a mutual belief that p. It can be proved that the two definitions are equivalent in the so-called Kripke structures (after Saul Kripke), the state space models that are traditionally used by logicians to model belief and knowledge.

The seminal work on the analysis of mutual belief is David Lewis's philosophical theory of *convention*. The economist Robert Aumann was the first to provide a mathematical characterization of a similar concept using set theory. More precisely, Aumann formalized the concept of *mutual knowledge* (also called *common knowledge*), which, contrary to mutual belief, has the property of being truthful. That is, while having the mutual belief that p is true does not necessarily imply that p is true, this is the case for mutual knowledge. More recently, formal accounts of mutual belief and mutual knowledge were proposed using the tools of epistemic logic by Ronald Fagin, Joseph Halpern, Yoram Moses, and Moshe Vardi; by Luc Lismont and Philippe Mongin; as well as by other logicians.

The concepts of mutual belief and mutual knowledge were widely employed to explain group activity, coordination, and communication. For instance, it was proved that mutual knowledge justifies the plausibility of equilibrium notions in game theory such as backward induction. An example of this is the well-known theorem proved by Aumann, which states the following: "For any non-degenerate game of perfect information in extensive form, mutual knowledge of rationality (i.e. the players have mutual knowledge that every player is an expected utility maximizer) implies the solution prescribed by backward induction," where backward induction is the process of reasoning backward in time, from the end of the game to the beginning of the game, to determine a sequence of optimal actions. Backward induction proceeds as follows. The first step is to determine the

optimal choice of the player who makes the last move of the game, where "optimal choice" means "a choice that maximizes the player's utility." The second step is then to turn to the second-to-last player and, taking the last player's choice as determined in the first step, to determine the optimal choice of the second-to-last player. And so on.

Moreover, mutual belief has been used to define the concept of "common ground" in a conversation, which is fundamental for discourse understanding and definite reference. Language use in conversation is a form of social activity that requires a certain level of coordination between what the speaker means and what the addressee understands the speaker to mean. Indeed, any utterance of the speaker is in principle ambiguous because the speaker could use it to express a variety of possible meanings. Common ground—as a mass of information and facts *mutually believed* by the speaker and the addressee—ensures coordination by disambiguating the meaning of the speaker's utterance. For example, suppose two different operas, *Don Giovanni* by Mozart and *Il Barbiere di Siviglia* by Rossini, are performed in the same evening at two different theaters. Mike goes to see *Don Giovanni* and the next morning sees Mary and asks, "Did you enjoy the opera last night?"—identifying the referent of the word *opera* as *Don Giovanni*. In order to be sure that Mary will take "opera" as referring to *Don Giovanni* and not to *Il Barbiere di Siviglia*, Mike has to believe that the night before Mary too went to see *Don Giovanni*, that Mary believes that Mike too went to see *Don Giovanni*, that Mary believes that Mike believes that Mary too went to see *Don Giovanni*, and so on.

Furthermore, common belief has been considered a fundamental constituent of joint activity and of shared and group intentions. For instance, according to Robert Bratman's well-known definition of shared intention, we can say that two agents share the intention of "painting the house together" only if they mutually believe that they both intend to paint the house together.

In computer science, mutual knowledge and mutual belief are central concepts in the analysis of properties of distributed systems. One of the results in that field is that mutual belief between two agents can only be attained if the communication channel is reliable (i.e., when an agent sends a message to another agent, then the latter will certainly receive the message, and the two agents have mutual belief of this), as noted by Ronald Fagin and associates.

Raimo Tuomela, Margaret Gilbert, and other philosophers working on social ontology have opposed mutual knowledge and mutual belief to other forms of collective attitude, such as the notion of *collective acceptance*. A property that clearly distinguishes mutual belief from collective acceptance is that mutual belief implies shared belief, while collective acceptance does not: When there is a mutual belief in G that p is true, then each agent in G individually believes that p, while it might be the case that there is a collective acceptance in G that p is true, and at the same time one or several agents in G do not individually believe that p is true. For example, the members of a parliament might collectively accept (qua members of the parliament) that launching a military action against another country is legitimate because by majority voting the parliament decided so, even though some of them—who voted against the military intervention—individually believe the contrary.

Emiliano Lorini

See also Collective Agents; Collective Goals; Collective Intentionality; Collective Rationality; Common Knowledge; Conventions, Logic of; Group Beliefs; Team Reasoning; We-Mode, Tuomela's Theory of

Further Readings

Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics, 4*(6), 1236–1239.

Aumann, R. J. (1995). Backward induction and common knowledge of rationality. *Games and Economic Behavior, 8,* 6–19.

Battigalli, P., & Bonanno, G. (1999). Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics, 53,* 149–225.

Bratman, M. E. (1992). Shared cooperative activity. *Philosophical Review, 101*(2), 327–341.

Clark, H. H., & Marshall, C. (1981). Definite reference and mutual knowledge. In A. Joshi, B. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge, England: Cambridge University Press.

Cubitt, R. P., & Sugden, R. (2003). Common knowledge, salience and convention: A reconstruction of David Lewis' game theory. *Economics and Philosophy, 19,* 175–210.

Fagin, R., Halpern, J., Moses, Y., & Vardi, M. (1995). *Reasoning about knowledge*. Cambridge: MIT Press.

Gilbert, M. (1989). *On social facts*. London, England: Routledge.

Heifetz, A. (1999). Iterative and fixed point common belief. *Journal of Philosophical Logic, 28*(1), 61–79.

Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Lismont, L., & Mongin, P. (1994). On the logic of common belief and common knowledge. *Theory and Decision, 37*, 75–106.

Lorini, E., Longin, D., Gaudou, B., & Herzig, A. (2009). The logic of acceptance: Grounding institutions on agents' attitudes. *Journal of Logic and Computation, 19*(6), 901–940.

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy, 25*, 701–721.

Tuomela, R. (1992). Group beliefs. *Synthese, 91*, 285–318.

Tuomela, R. (2002). *The philosophy of social practices: A collective acceptance view*. Cambridge, England: Cambridge University Press.

Narrative in Historical Explanation

The topic of narrative in historical explanation excites controversy because of deep disagreements regarding what makes such narratives genuinely explanatory. This entry reviews two areas of philosophical controversy regarding the status of narrative in historical explanation: epistemological and metaphysical issues.

In this context, "narrative" connotes a presentation in a story-like fashion of presumably related events that result in a particular outcome. Two sets of philosophical issues have dominated debate about narratives qua explanations. One concerns formal and epistemological issues, the other metaphysical ones. (1) Formal and epistemological concerns typically ask after theoretical articulations of assumed connections and so impose justificatory demands on relationships cited as explanatory. (2) A fundamental metaphysical issue with regard to narrative in historical explanations centers on narratives per se as a form of representation of the reality of the past.

1. Regarding the first set, a standing concern has been that what narratives often suggest as *causes* involve un- or undertheorized connections, such as *reasons* of various sorts. Relatedly, a narrative might appear to explain by providing information unique to a period and its thought. But then, the very reasons presented as explanatory seemingly count against integrating such an explanation with those generalizing and integrative models of explanation familiar from the natural sciences. So questions remain about how narratives explain.

In the heyday of logical positivism, when an assumption regarding a common logical form to properly scientific explanations reigned supreme, narratives were explanations in some derivative sense, one that depends on a received account of scientific explanation. Formally, narratives were stigmatized as at best "explanation sketches" (a term used by Ernest Nagel). To vary a bit C. G. Hempel's famous example, suppose that someone narrates that the water in his car radiator froze and then the radiator burst. The person concludes by saying, "I forgot to put in antifreeze." But this brief story only explains because it can be "converted" into an account that justifies the connections—for example, the temperature at which water freezes and the tensile strength of the radiator. By contrast, when Raul Hilberg summarizes a key thesis of his magisterial narrative of the Holocaust, *The Destruction of the European Jews,* he pointedly denies that acts eventuating in the creation of death camps followed from one another in any predictable or preordained way. Rather, such a narrative explanation can only chart, in Karl Schleunes's phrase, the "twisted road to Auschwitz." This typifies a historian's sense of a narrative, at least insofar as it denies that a narratively presented sequencing of events even possibly "converts" to some more generalizing form of explanation such as those found in the natural sciences.

Questions regarding form—what elements a proper narrative explanation must contain—and epistemology—what theory underwrites the imputed causes—generated a division of views that persists

unabated. Epistemologically, some defend narrative explanations as a distinctive format for teleological or purposive explanations. Another epistemological approach construes narratives as a sui generis form of explanation. On this view, narratives either represent a special cognitive act, as Louis Mink holds, or achieve their "explanatory effect" because, in a view famously advocated by Hayden White, such story forms play a special role in our cognitive economy.

Mink's influential account emphasizes narrative as a mode of comprehension that differs categorically from the type of explanations that the natural sciences offer. Mink's position can be distinguished from those who, following Hayden White, emphasize narrative as an imposition of order by means of cultural poetics. For White, the writing of history should be understood as just a special case of literary creation and thus should be analyzed in that spirit. Mink, in contrast, emphasizes that the key formal characteristic of historical narrative as a type of explanation resides in the fact that it consists of a special form of *retrospective* of sense making. For an important feature about knowledge of the past, as Arthur Danto's work establishes, involves the fact that statements true of the past could not be known as true (or known at all) at that time. Danto's canonical example is "The Thirty Years War began in 1618." Although now true of what happened in 1618, this statement became true of events in 1618 some 30 years after the fact. What happens later creates truths of times past that yet were not true or knowable at just those times. And while Mink's position does not exclude consideration of cultural poetics, its emphasis falls on historical narratives as retrospectively fashioned accounts of events already lived.

2. A second currently debated question flows directly from the formal considerations urged by White and by Mink and parallels the realism/anti-realism debate in the philosophy of science, at least in the following respect. For both White and Mink emphasize histories as human fashionings—that is, narratives as *constituting* historical events. But the question then arises whether this implies that narratives per se (and not just their individual statements of fact) cannot be true, in the sense of correctly representing *the* past. (Use of the definite article is key here.) What happens to the realist intuition that narrative histories could possibly represent the past *wie es eigentlich gewesen* ("as it really was")?

Indeed, such metaphysical/representational questions about historical narrative explanations have now come to the fore in philosophical debates about the reality of the past and narrative as a form of representation of this assumed reality. Mink anticipated the unsettling implications of what happens to an understanding of human history in the absence of a belief in any "master narrative" or "universal history." Once a belief in a universal history goes, no single narrative can claim metaphysical precedence. As a result, history cannot be imputed a determinate structure perceivable *sub specie aeternitatis*. In sum, debates about narrative in historical explanation begin as ones of form or epistemology. But the sharpest philosophical challenge that emerges from these discussions has been to the metaphysical assumptions underlying belief in the reality of the past—in other words, to the idea that human history has any independent determinate form.

Paul A. Roth

See also Causation in the Social Sciences; Explanation, Theories of; Explanation Versus Understanding; Historicism; *Naturwissenschaften* Versus *Geisteswissenschaften;* Philosophy of History; Realism and Anti-Realism in the Social Sciences

Further Readings

Danto, A. C. (1985). *Narration and knowledge.* New York, NY: Columbia University Press.

Mink, L. O. (1987). *Historical understanding.* Ithaca, NY: Cornell University Press.

Roberts, G. (Ed.) (2001). *The history and narrative reader.* New York, NY: Routledge.

Roth, P. A. (2002). Ways of pastmaking. *History of the Human Sciences, 15,* 125–143.

White, H. (1987). *The content of the form: Narrative discourse and historical representation.* Baltimore, MD: Johns Hopkins University Press.

Naturalism in Social Science

Naturalism in the social sciences—namely, that the subject matter of the social sciences should be seen as an element of the natural world and could be explained accordingly—has an ontological face and a methodological face: that social institutions

or human phenomena in general can be reducible to natural ones or that they are explainable naturalistically. This entry discusses critically each version of naturalism in turn.

Ontological Naturalism

Ontological naturalism holds that the objects and phenomena of interest in the social sciences are part of, or aspects of, the natural world. As biological entities and phenomena are part of, or aspects of, the physical world, so also is the subject matter of the social sciences a part of the natural world. Culture is natural for humans. The parallel with biological phenomena is instructive for the naturalist. Biological systems are relatively complex physical systems. They have interesting and somewhat distinctive properties. Examples include the capacity for reproduction, for certain forms of homeostasis, and for assimilation of nutrients. Such capacities, or properties, are commonly not instanced in the nonbiological components of the physical world. They make for a distinctive biological subject matter—though clearly it remains a dependent part of the physical world. Thus, a biological entity's capacity for self-replication is based in (or is reducible to) the simpler dispositional capacities of complex molecules and ultimately in the behavior of sequences of atoms in molecules. Plants are a class of system with the capacity for converting incident solar energy into stored chemical energy. Living things generally have the capacity for reproduction. In all such matters, there is no suggestion of anything "supernatural"—no suggestion that there might be some force that eludes physics. This is not to suggest that such systems can only be understood at the level of physics—for typically, they can be understood in terms of regularities involving biological features or properties. It is to say that they are ultimately complex physical systems, and thus natural systems.

Ontological naturalism extends this picture to psychological phenomena and to human groups and group phenomena. These are yet more complex phenomena within the natural (and physical) world. Again, these phenomena may turn on features and regularities that are not features of the simpler physical/natural systems at lower levels of organization. There is no transcription, and no transcription errors, mutation, mutation repair, and the like, within atoms—as these are features or properties of cells and cellular phenomena. There are no beliefs within atoms, as there are within humans or like creatures. There is no norm formation and conformity and no elections and victories within atoms—just within various social groups. Yet such higher-order phenomena and features plausibly arise out of organized lower-order phenomena and properties. To capture the resulting dependency of the higher order on a basis in the natural, ultimately physical, world of which it is then a composed part, many ontological naturalists write of a *supervenience* of the one on the other. A set of properties, *S,* is said to supervene on another set of properties, *B,* just in case were a system alike with respect to the *B*-properties, it would be alike with respect to the *S* properties (how things are with respect to the *B*-properties determine how things are with respect to the *S* properties).

Properties with a historical aspect seem to call for special comment. For example, on the phylogenetic species concept, a species is understood as a location on the phylogenetic (evolutionary) tree. On this concept, a molecule-for-molecule parallel of a population of *Canis lupus* engineered in a lab (and thus not resulting from the standard evolutionary mechanism) would not count as a population of *Canis lupus*. However, it would remain a part of the natural world—relevantly understood. Similarly, duplicates of U.S. Treasury notes are not U.S. Treasury notes. Yet the notes, and the various human institutions, are here thought of as a part of the natural world. (Think also of various geological epochs or periods—the Carboniferous, for example—and historical periods—the European Middle Ages. While historically specific, each one is a part of the natural world.)

Methodological Naturalism

Methodological naturalism is the position that, as a part of the natural world, social phenomena are best studied using methods that have borne fruit in the study of the rest of the natural world.

Such study has commonly turned on careful attention to regularities—to generalizations that can be refined under test and that can be deployed to understand the patterns of counterfactual dependencies holding with respect to particular episodes or classes of events. The systems, or classes of cases so understood, commonly are local and limited—so that the relevant generalizations would also be limited. That there are such systems—for example, in ecology—is

a contingent matter. Ecosystems (or more specifically tundra or limnological ecosystems) are local and contingent—that is, systems thrown up by the contingencies of evolution within a limited planetary system. However, there are generalizations that hold for such systems. Ecology advances by coming to understand such regularities in terms of generalizations with a limited degree of invariance. Often, the understanding one has of such natural systems comes by way of deploying generalizations treating interrelated processes at various levels. Ecology has fairly weak generalizations to call its own but borrows heavily on generalizations from a range of related fields.

Methodological naturalists project parallel developments within fruitful social science. For example, suppose one sought to understand the prevalence of certain social norms in a certain cultural context. According to the methodological naturalist, one will be in a position to do so when one has developed and empirically refined a range of generalizations. Plausibly, some generalizations might have to do with the social transmission of cultural ideas—perhaps generalizations treating a replicator dynamics having limited analogies with genetics (as in evolutionary approaches to culture). Some generalizations might have to do with human psychology, particularly generalizations having to do with characteristic human social preference. Some generalizations might originate within general cognitive psychology and suggest ways in which norm representations might be stored and activated (as accounts of the dynamics of social norms suggest). In all cases, the generalization in question should be subject to refinement either in the lab or in other experimental settings, and with careful cross-cultural investigation.

While ontological naturalism has few natural enemies today, methodological naturalism is opposed by a form of methodological separatism that insists (a) that social phenomena are best appreciated by way of "interpretation" and (b) that the interpretation operates by a kind of logic that is unlike any in the natural sciences. This is reflected in the *hermeneutical* tradition. A recent, distinct version of interpretivism, originating from within analytic philosophy, takes inspiration from Donald Davidson's views on interpretation: He understands interpretation as essentially conditioned by a normative—nondescriptive—presumption of rationality. Some find this line of thought untenable, while others advance a moderate naturalist account of interpretation.

David K. Henderson

See also Biology and the Social Sciences; Causes Versus Reasons in Action Explanation; Explanation Versus Understanding; *Naturwissenschaften* Versus *Geisteswissenschaften;* Normativism Versus Realism; Reductionism in the Social Sciences; Social Ontology, Recent Theories of; Supervenience

Further Readings

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. New York, NY: Cambridge University Press.

Camerer, C., & Fehr, E. (2004). Measuring social norms and social preferences using experimental games: A guide for social scientists. In J. Henrich, R. Boyd, S. Bowles, & C. Camerer (Eds.), *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies* (pp. 55–95). Oxford, England: Oxford University Press.

Darden, L., & Maull, N. (1977). Interfield theories. *Philosophy of Science, 44,* 43–64.

Davidson, D. (2001). *Inquiries into truth and interpretation* (2nd ed.). Oxford, England: Clarendon Press.

Henderson, D. (1993). *Interpretation and explanation in the human sciences* (SUNY Series in the Philosophy of the Social Sciences). Albany: SUNY Press.

Henrich, J. (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology, 208*(1), 79–89.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H. (Eds.). (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford, England: Oxford University Press.

Rosenberg, A. (2008). *Philosophy of social science* (3rd ed.). Boulder, CO: Westview Press.

Woodward, J. (2000). Explanation and invariance in the special sciences. *British Journal for the Philosophy of Science, 51*(2), 197–254.

Naturalized Epistemology

Naturalism in its 20th-century American incarnation seeks to align philosophical inquiry with scientific inquiry generally. In this form, naturalism rejects a priori claims to knowledge (i.e., it rejects a view of knowledge as not explicable as part of the natural world) and links justification to the methods of science, broadly understood. Unlike positivism, naturalism offers no rigid demarcation criterion

regarding what defines a science. In post-positivism, questions arise on how to delineate what counts as a science, and this creates a lingering strain with naturalism as originally conceived.

Since much of the discussion of naturalized epistemology emanates from Willard Van Orman Quine's landmark essay on the topic, debate on this topic might be presumed to also inherit Quine's holism with regard to theories and his extremely liberal notion of what constitutes a science. But this has not turned out to be the case. Broadly specified in a Quinean spirit, naturalized epistemology could incorporate methodologically much beyond any list of methods native to the natural sciences and so includes those of, for example, any of the social sciences and history. As a naturalist, Quine assumes no fixed methodological essence that serves to demarcate disciplinary kinds. Read less in the spirit of American naturalism, naturalized epistemology can be taken to imply a reductionist program, with the natural sciences supplying the licit elements of any proposed reduction. But inasmuch as such readings beg the question of how to establish what to count as a science, they will not be considered further here.

Disagreements regarding what the term *naturalized* implies in consequence typically turn on ambiguities regarding what a science is, since explications of "naturalism" invoke some notion of science. It might be thought that agreement on the purview of the notion of epistemology would prove unexceptional no matter how contentious the interpretation given to the modifier "naturalized." But it turns out that using this modifier generates uncertainty with respect to epistemology so modified. Even supposing that everyone agrees that epistemology covers the philosophical analysis of knowledge, such apparent agreement would only mask a core philosophical controversy. For what knowledge is, and so what epistemology in fact examines and encompasses, will depend on how the term *naturalized* functions to modify *epistemology*. Thus, the topic of naturalized epistemology elicits controversy not only with regard to what to accept as a *naturalized* epistemology but also as to how to construe "epistemology" so modified. In short, determining what sets the scope of the term *knowledge* has emerged as a key point of contention with respect to providing any account of naturalized epistemology.

The "Autonomy of Knowledge" View

Two antithetical answers to the question of what knowledge is dominate contemporary epistemology. One answer, claiming a heritage to Thomas Reid and G. E. Moore, premises its approach to epistemology and a determination of its subject matter on what could be termed the "autonomy of knowledge" assumption. Knowledge so conceived constituted a domain prior to philosophy ever arriving on the scene. On this assumption, there exists a pretheoretic ("commonsense") notion of knowledge that any analysis of knowledge must respect. From this perspective, tests of philosophical analyses of knowledge then proceed by holding them hostage to "intuitions" regarding a fit between these analyses and a prior sense of what belongs in the domain.

On this "autonomy of knowledge" view, then, epistemology begins with an established subject matter—some pretheoretically ascertained domain of the known—prior to, and independently of, modifiers such as "naturalized." The modifier thus only signals a particular approach with respect to an independently established category, as in "Italian cuisine." But the category has independence and a breadth apart from its modifier.

Further, by assuming the "autonomy of knowledge" view of epistemology, any efforts to bring science to bear on knowledge claims will be circumscribed by knowledge pretheoretically understood. People pretheoretically have a great deal of knowledge, and they always already know (more or less) what propositions count as instances of knowledge. Philosophical analysis contributes just self-awareness in the form of a specification of the conditions (i.e., the analysis) of knowing. The object of analysis—knowledge—exists prior to, and independently of, any science. From this perspective, a naturalized epistemology only contributes whatever resources the sciences might offer that abet this task of making explicit already existing conditions of knowledge. Science then plays the handmaiden to philosophy.

The "Autonomy of Naturalized Epistemology" View

A second and very different view of naturalized epistemology emerges, however, on the assumption that pretheoretic views have no particular standing to determine or adjudicate what passes for knowledge. On what may be termed the "autonomy of naturalized epistemology" assumption, whatever falls within the scope of a naturalizing process determines the domain of epistemology and so defines knowledge. Rather than assuming, that is, an autonomous domain of knowledge before philosophy comes on

the scene, this approach to naturalized epistemology defines knowledge in naturalistic terms and proceeds from there. Here, the subject matter of epistemology can only be determined in conjunction with what current sciences endorse. People of course belong to linguistic communities that already use terms such as *knowledge, justification, good reasons,* and so on. But once in place, science sets the standard for what passes as knowledge. So even on the assumption that science evolves from ordinary understanding, the sciences ultimately come to constitute the domain of knowledge properly so called.

Quine argues for this view specifically, maintaining that with the collapse of foundational programs epistemology loses any standing to prescribe to science a standard of knowledge apart from whatever the sciences collectively endorse. For in the absence of foundationalism, there exists no extrascientific gold standard for adjudicating knowledge claims. And with the standards of the various sciences in hand, it can then be decided which candidates for knowledge meet them. Since standards come bundled with theories generally, knowledge then consists of all that these theories imply.

Framing the Debate

Although it might seem as if intermediate positions between these two extremes—the "autonomy of knowledge" and the "autonomy of naturalized epistemology"—should be possible, further consideration indicates otherwise. The chief sticking point arises with regard to what will count, in any approach, as knowledge. Either there exists an identified nonscientific standard of knowledge, or there does not. Naturalizers say that there does not. A consistent naturalizer cannot canonize both common sense and science; for that would be to certify as epistemologically appropriate methods not licensed by science. Nonnaturalizers hold to their Moorean sensibilities and so assign philosophical primacy to a pretheoretically determined knowledge domain.

Either pretheoretic intuitions about knowledge trump science or vice versa. For a nonnaturalist, invoking science as an aid begs all the critical philosophical questions. Assigning science an epistemological role must appear inexplicable or arbitrary. For if science sets a (rather than "the") standard of justification, what reveals this? Certainly, nothing related to "common sense" or pretheoretic intuitions about knowledge. So science on the nonnaturalist view awaits justification; it cannot plausibly function as a touchstone, however modest, of knowledge. Conversely, to endorse science would be to endorse a standard with no ultimate necessary connection to whatever informs commonsense intuitions about knowledge. Philosophically, one cannot have it both ways. Thus, as noted above, one will be pushed to the extremes represented by one or the other of the autonomy views.

Two red herrings often further cloud the debate with regard to naturalized epistemology. One, already mentioned, insists that a naturalized epistemology entails a reductionist program, with the terms of reduction having to come from the physical sciences. While it is not clear whether or not positivism even in its heyday mandated such an approach to epistemology, post-positivism, no such implication can be said to follow from naturalism. In the absence of a demarcation criterion, what to call a science can plausibly be said to include all disciplines that seek to provide an integrated and systematic account of their subject matter. A naturalized epistemology uses all the tools of the various sciences to study how the subject matters of these sciences (and so what can pass for knowledge) come to be, including how those sciences themselves as theories of these subject matters come to be. No necessary or sufficient conditions exist for separating sciences from nonsciences, but this hardly matters. Indeed, the disunity of science has emerged as a working hypothesis for understanding the sciences as we find them. Inevitably, a naturalized epistemology will depend on what a society recognizes as legitimate science.

A second red herring issue that also often surfaces in debates with regard to naturalized epistemology concerns the role of so-called normative factors with respect to an analysis of knowledge, for example, what makes for the goodness of reasons. The charge goes that science, and so by extension any naturalized epistemology, can only focus on descriptive/causal aspects of cognition. Nothing in science permits an analysis of evaluative practices, for evaluative practices can never be caught in a descriptive/causal net; they are items of another sort. Hence, according to this criticism, the very term *naturalized epistemology* constitutes an oxymoron; to the extent that the subject matter allows of naturalization, to that extent it cannot address the primarily normative concerns of epistemology.

Of course, those making this criticism cannot deny that many social sciences study the genesis,

evolution, and change of norms within groups. The complaint only has a point if the norms at issue cannot be identified with those "transitory" or "merely social" norms that have their lineage analyzed in this way. The investigation then must be in primary part conceptual, not empirical, since it asks after the appropriateness or rightness of a type of judgment, not how this or that rule came as a matter of fact to be taken for a norm. Only a mode of philosophy not fully or at all naturalized stands ready to undertake such inherently conceptual investigations.

Here, the debate seems to end. A naturalist can only dig in her heels and question why it must be the case that any norm has a more exalted status than whatever reasons led a group in the first place to adopt it. Does this imply that all norms receive a functional explication? No. It merely points to the fact that disciplines tend to settle over time what their norms are. These can be studied, explained, and, yes, critiqued by examining through various means—historical, sociological, economic, and so on—how norms come to be and pass away.

Naturalized epistemology denies what traditional philosophy took it to be the purpose of first philosophy to provide, namely, a metaperspective from which to adjudicate the "really best" from among those norms that wash up on a society's shores in the ebb and flow of history. But a naturalized epistemology offers no such promise of an atemporal or absolute perspective on cognitive standards. A naturalized epistemology can help one learn how what now passes for knowledge came to so pass, and by doing so possibly suggest what may have been problematic in that process. In this clear respect, a naturalized epistemology, qua epistemology, does contribute to the eternal philosophical project of increasing self-understanding. But *qua naturalized* epistemology, it promises no more than that, that is, a bringing to self-awareness through science how beings like us come to possess the sciences that we do.

Paul A. Roth

See also A Priori and A Posteriori; Duhem-Quine Thesis and the Social Sciences; Epistemology; Ethno-Epistemology; Feminist Epistemology; Naturalism in Social Science; *Naturwissenschaften* Versus *Geisteswissenschaften;* Normativism Versus Realism; Pragmatism and the Social Sciences; Reductionism in the Social Sciences; Social Epistemology

Further Readings

Danto, A. C. (1967). Naturalism. In P. Edwards (Ed.), *The encyclopedia of philosophy* (Vol. 5, pp. 448–450). New York, NY: Collier Macmillan.

Kim, J. (1988). What is "naturalized epistemology"? *Philosophical Perspectives, 2,* 381–405.

Kitcher, P. (1992). The naturalists return. *Philosophical Review, 101,* 53–114.

Kornblith, H. (Ed.). (1994). *Naturalizing epistemology* (2nd ed.). Cambridge: MIT Press.

Laudan, L. (1990). Normative naturalism. *Philosophy of Science, 57*(1), 44–59.

Quine, W. V. (1969). Epistemology naturalized. In W. V. Quine (Ed.), *Ontological relativity and other essays* (pp. 69–90). New York, NY: Columbia University Press.

Roth, P. A. (2006). Naturalism without fears. In S. Turner & M. Risjord (Eds.), *Handbook of the philosophy of science: Vol. 15. Philosophy of anthropology and sociology* (pp. 683–708). Boston, MA: Elsevier.

Stroud, B. (1996). The charm of naturalism. *Proceedings and Addresses of the American Philosophical Association, 70*(2), 43–55.

Turner, S. P. (2010). *Explaining the normative.* Malden, MA: Polity Press.

Zammito, J. (2004). *A nice derangement of epistemes: Post-positivism in the study of science from Quine to Latour.* Chicago, IL: University of Chicago Press.

Naturwissenschaften Versus *Geisteswissenschaften*

This entry presents the celebrated contrast between the natural and the human sciences as it developed historically in the 19th century around the core ideas of Wilhem Dilthey. The entry focuses on Dilthey's thought and discusses the various epistemological theses characterizing the distinction between the sciences of nature and the human sciences. It also presents the Neo-Kantian distinction between "nomothetic" and "idiographic" methods.

Introduction

The distinction in German between the *Naturwissenschaften,* or natural sciences, and the *Geisteswissenschaften* received its classical definition in the writings of the German philosopher and historian Wilhelm Dilthey (1833–1911) and has continued to shape our thinking about the sciences. The term *Geisteswissenschaften* is difficult to translate.

The literal translation "sciences of spirit" points back to the Hegelian view of philosophy as the one and only science of spirit (*Wissenschaft des Geistes*). Dilthey, however, conceives the *Geisteswissenschaften* pluralistically and in empirical terms, which makes the literal translation misleading. *Geisteswissenschaften* is therefore commonly and more appropriately translated as "human studies" or "human sciences."

Origins and Legacy

The plural use of *Geisteswissenschaften* already occurred in the writings of the Romantic natural philosopher Lorenz Oken in 1817 and the historian Johann Gustav Droysen in 1843. But it was not until it was used to translate John Stuart Mill's use of the term *moral sciences* in 1849 that the *Geisteswissenchaften* began to be conceived more empirically and in terms of the methods that are appropriate to them. This was then programmatically worked out by Dilthey in his *Introduction to the Human Sciences* of 1883 and his *Formation of the Historical World in the Human Sciences* of 1910.

The human sciences include both the humanities and the social sciences. For that reason, the twofold *Natur-Geisteswissenschaften* distinction found in German universities is now also expanded into the threefold *Natur- Geistes-und Sozialwissenschaften* distinction. One of the important contributions of the twofold distinction is that it brings reflective and disciplinary considerations to bear on the humanities by not sharply separating them from the social sciences. Social scientists have been more eager to emulate the explanative and statistical methods of the natural sciences. The more inclusive notion of the human sciences has led to the development of methods of *understanding* and *interpretation* that are applicable to both more humanistic disciplines such as literary history and more systematic disciplines such as sociology. The sociologist Max Weber (1864–1920), for instance, claimed that it is necessary to understand the Protestant work ethic to explain why capitalism first emerged in countries such as The Netherlands, England, and the United States. Weber and philosophers such as G. H. von Wright (1916–2003) consider *understanding* a preliminary to *explanation*. Dilthey and many phenomenologists, however, conceive understanding to be more comprehensive than explanation.

Wilhem Dilthey

Dilthey's *Introduction to the Human Sciences* examines the relation between the natural and human sciences in two ways. The first way is to consider how the human sciences supplement the natural sciences and to what extent their goals differ. This is the general approach of Book One. The second approach takes stock of the way the human sciences have emerged from human self-reflection and is most fully developed in the drafts for Books Four to Six. What Dilthey calls self-reflection or anthropological reflection is both theoretical and practical and establishes its own parameters.

These two approaches to the sciences are developed in the following two sections; the third section deals with subsequent developments.

Human Sciences and Natural Sciences

In Book One of the *Introduction to the Human Sciences,* Dilthey differentiates the human sciences from the better-established natural sciences. He points out that whereas the natural sciences provide causal explanations of the phenomena of outer experience, the human sciences attempt to understand the *meaning* of both inner and outer experience. Acknowledging that mental processes are influenced by physical processes, he nevertheless argues that this involves a functional rather than a causal relation, which entails that mental life can have a relative independence. There are aspects of the life of the human spirit that cannot be explained by the natural sciences, and throughout human history different disciplines have arisen to deal with the issues faced by human beings in their social and cultural life. According to Dilthey, the human sciences, as they have evolved over time, contain three classes of assertions. One class describes reality as perceived, the second class explicates the uniform behavior of partial contents of this reality, and the third class expresses value judgments and prescribes rules. In the human sciences, descriptive-historical accounts of what is singular and individual are as important as the explication of theoretical uniformities. And both are governed by a normative system that connects values, ideals, and the aim to shape the future.

This interdependence of the theoretical and practical in the human sciences makes them less hierarchical than the natural sciences. Initially, Dilthey

did consider psychology to be the first of the human sciences, similar to how physics had become the foundational explanative natural science. But he always made it clear that psychology can only serve an analogous role if it stays within the limits of a descriptive discipline and renounces the hypotheses of explanative psychology about the underlying elements of mental life, most of which cannot be tested. Descriptions of psychic processes can delineate general structural regularities, but they cannot be abstracted from the rest of reality. Thus, a psychological trait such as thrift cannot be understood apart from the economic and social context within which individuals find themselves. Later, as Dilthey's thought became more hermeneutical, psychology would be stripped of even this limited foundational role. But he never doubted that psychology had a place among the human sciences and their efforts to understand the spiritual aspirations that are manifested in human history.

The Baden School of Neo-Kantians: Reaction to Dilthey

Dilthey's claim that psychology is a human science was rejected by the Baden-school Neo-Kantians Wilhelm Windelband (1848–1916) and Heinrich Rickert (1863–1936), who considered it to be a natural science. Insisting on a sharp distinction between psychological and epistemological accounts of consciousness, they restricted psychology to the search for empirical causal laws linking physiological and psychological states. Only the transcendental standpoint of epistemology can disclose the spontaneity of consciousness. And to do justice to the historical developments of the human spirit, they turn to what they call the cultural sciences (*Kulturwissenschaften*). However, Marburg Neo-Kantians, like Ernst Cassirer (1874–1945), were willing to speak of both cultural and human sciences.

Windelband attacked Dilthey's distinction of the natural and human sciences for focusing too much on differences of subject matter and proposed that the formal distinction between *nomothetic* and *idiographic* methods is more appropriate. According to Windelband the natural sciences are nomothetic in developing lawful uniformities, and the cultural sciences are idiographic in focusing on unique historical patterns. Dilthey rejects this sharp division as invalid for both kinds of science. The natural sciences are not exclusively nomothetic. He points out that biology involves both a general theory of animal functions and a descriptive classification of animals. Moreover, there are many human sciences, such as economics and sociology, that have established uniformities within certain systems of interaction and cooperation that have come to be differentiated and institutionalized over the course of historical life. The understanding of individuality that characterizes the human sciences is not merely the appreciation of ideographic singularity but requires that we consider the uniformities that provide the background for individuation. These uniformities may be causal or structural in nature.

Lived Experience and Self-Reflection

In Book Four of the *Introduction to the Human Sciences,* which was published posthumously, Dilthey observes that everything real is a fact of consciousness and subject to the conditions of consciousness. Reality is already there for us or present before it is represented cognitively or scientifically. Our lived experience is reflexively given to us in a way that precedes psychological description and the theoretical analysis of epistemology. For reflexive awareness, which is prereflective, there is no distinct self to be described. Only gradually does the continuum of lived experience differentiate itself into inner and outer experience, a reflective self and an objective world. And when the overall nexus of lived experience is reflected upon, it provides the context for both theory and practice. Dilthey makes "self-reflection" about this context, not merely epistemology, the foundation of all the sciences. Reflection's most basic task is to analyze what is given and to relate the partial contents of experience in accordance with the forms and rules of logic. The human sciences must analyze the givens of experience without losing sight of their larger life context, and experience itself must be probed for the ways we engage the world in terms of thought, feeling, and volition. The natural sciences analyze the givens of experience precisely to the extent that we do not contribute to them. Their results are thus more abstract and primarily theoretical. They deal with fewer variables and can thus be more successful in testing explanative hypotheses.

From this perspective, the natural sciences are more derivative than the human sciences. To be sure, the understanding aimed at by the human sciences

must incorporate the empirical findings and causal uniformities of the natural sciences when describing human behavior and historical events. But it is the *meaning context* defined by the human sciences that guides the judgment on where to look for causal conditions relevant to human behavior. Similarly, the goals set by human agents codetermine how nature as a system of means applies.

Dilthey argued for a radically pluralistic theory of the human sciences. Just as there cannot be an explanative foundational human science such as physics, there should not be a monolithic, overarching discipline like Hegel's philosophy of history or Auguste Comte's sociology. There are no laws of history at large, and sociology cannot do justice to all aspects of human interaction. History can only be understood by analyzing it into the specific systems of interaction that have developed over time. Dilthey divides them into what he calls voluntary productive systems of cooperation, such as economic and political organizations, and the institutional formations in which individuals find themselves. It is only within these more defined sociocultural systems that we can expect to arrive at uniformities and laws.

Further Developments

Although nature has become alien to us since the rise of modern science and must be explained in hypothetical intellectual terms, Dilthey thinks that much of our psychosocial life can be understood nonhypothetically from within. We participate in the life of history, which means that its understanding (*Verstehen*) engages not only our intellect—what Immanuel Kant had called the faculty of understanding (*Verstand)*—but all the cognitive, affective, and volitional powers of the mind. In his "Ideas for a Descriptive and Analytic Psychology" (1894), Dilthey asserts that the inner connectedness of lived experience already provided an indeterminate understanding, which could then be further explained in detail by means of specific and testable hypotheses. But in the essay "The Rise of Hermeneutics" (1900), he comes to the realization that the inner intelligibility of lived experience does not yet constitute understanding. The way we express ourselves, whether in communication or in action, is a crucial intermediary in defining ourselves. Understanding can only be reliable if it proceeds through the interpretation of human objectifications.

In the *Formation of the Historical World in the Human Sciences* (1910), Dilthey provides a final delineation of the relation between the natural and human sciences. They cannot be distinguished merely as two spheres of facts. They often treat the same facts, such as language: one in terms of the physiology of the speech organs, the other in terms of syntax and semantics. But it would be a mistake to think that the natural sciences deal with the external aspects of language and the human sciences with the inner core. The human scientific approach to language must take into account not only the physiological conditions of speech formation but also the physical manifestations of the way language is used. This leads Dilthey to point to the physicality of linguistic expressions as a means of understanding. It is through the interpretive path from without to within that understanding begins to define itself. Everything that confronts us in the human historical world—whether it be an individual linguistic expression, a human action or deed, a social interaction, or the practice of a public institution—is an objectification that must be referred back from its outer, sensory aspect to one that is withdrawn from the senses and therefore inner. This inner aspect is in the first instance not mental but normative. Thus, what we see happening in the courtroom, where criminals are brought to justice, is the expression of a purposive system of legal norms and functional rules. Here, interpretation finds an inner meaning that is not some psychical content but points to a historical formation that has its inherent structure and lawfulness. Even in the case of an individual creation such as a novel, what is expressed is not some complex of the novelist's mental states but a structural nexus consisting of motifs, plot, and characters created on the basis of those states. The understanding of this literary nexus need not appeal to psychology except if there are sharp deviations from the conventional ways in which characters are portrayed or shown to interact.

The natural sciences relate the objects of experience to each other and thus constitute an external natural system. The human sciences relate those objects that can be considered human objectifications to historical systems that can be called "inner," not because they belong to us but because we participate

in them. The understanding of these inner structures is always indirect and requires interpretation.

Concluding Remarks

Dilthey contrasted the human sciences to a classical, law-based conception of natural science. Today, the natural sciences show many interpretive features as well, which has led some to speak of a universal hermeneutics. But interpretation is only hermeneutical if meaning is assessed normatively in relation to values and purposes.

Rudolf A. Makkreel

See also Explanation Versus Understanding; Hermeneutics, Phenomenology, and Meaning; Neo-Kantianism; Phenomenological Schools of Psychology; Philosophy of History; Weber and Social Science: Methodological Precepts; Weber's *Verstehende* Approach

Further Readings

Dilthey, W. (1989). *Wilhelm Dilthey—Selected works: Vol. 1. Introduction to the human sciences* (R. A. Makkreel & F. Rodi, Eds.). Princeton, NJ: Princeton University Press. (Original work published 1883 and 1982)

Dilthey, W. (2002). *Wilhelm Dilthey—Selected works: Vol. 3. The formation of the historical world in the human sciences*. Princeton, NJ: Princeton University Press. (Original work published 1910)

Dilthey, W. (2010). *Wilhelm Dilthey—Selected works: Vol. 2. Understanding the human world*. Princeton, NJ: Princeton University Press. (Original work published in the 1890s)

Lessing, H.-U., Makkreel, R., & Pozzo, R. (2011). *Recent contributions to Dilthey's philosophy of the human sciences*. Stuttgart, Germany: Frommann-Holzboog.

Makkreel, R. (1992). *Dilthey: Philosopher of the human studies* (2nd ed.). Princeton, NJ: Princeton University Press.

Rickert, H. (1986). *The limits of concept formation in natural science: A logical introduction to the historical sciences*. Cambridge, England: Cambridge University Press. (Original work published 1902)

Von Wright, G. H. (1971). *Explanation and understanding*. Ithaca, NY: Cornell University Press.

Weber, M. (2009). *The Protestant ethic and the spirit of capitalism* (T. Parsons, Trans.). New York, NY: W. W. Norton. (Original work published 1904–1905)

Windelband, W. (1924). Geschichte und naturwissenschaft [History and natural science.] In *Präludien* (Vol. 2). Tübingen, Germany: J. C. B. Mohr.

Neo-Kantianism

Neo-Kantianism broadly refers to a philosophical movement mainly in Germany and France beginning in the second half of the 19th century, peaking between 1880 and 1910 and declining after World War I. As the name of this movement suggests, the members of this movement intended to revive the spirit (not the letter) of Immanuel Kant's philosophy in various manners.

This adherence to Kant is remarkable, for three reasons. First, different philosophers picked out different, and in part contradicting, moments from the Kantian oeuvre, so as to make it nearly impossible to indicate one main tendency of this broad movement, other than the general label of "philosophizing in the Kantian spirit." Second, Kant was hardly the only philosopher who was invoked in this movement; other theorists of influence for the Neo-Kantians were Johann Gottlieb Fichte and even G. W. F. Hegel. Third, some thinkers of crucial influence, especially in the early phase, were not philosophers but scientists who were pushed to more general philosophical musings through their experiments and discoveries in laboratories and found, to their surprise, an ally in Kant. Ironically, it was the scientistic and positivistic spirit of the time that renewed the urge on the part of many scientists to return to the sober spirit of Kant and also to renew a new sense of idealism or criticism. The close connection to cutting-edge science, however, remained a staple of Neo-Kantian practice.

Reception and Recent Revival

Before going into the philosophical details of the movement, it needs to be emphasized that, despite their falling into near-total oblivion by the second half of the 20th century, the Neo-Kantians were the most influential and most highly regarded thinkers in Germany and France around 1900, which has led Jürgen Habermas to characterize their stance in

academia and culture at large at the time as "imperial." This also explains why the philosophers to critically reject them—most notably the members of the emerging phenomenological movement and the Vienna Circle—bashed the movement of Neo-Kantianism as a whole with such force. It seemed that no progress could be made in philosophy without rejecting wholesale the philosophical establishment of the time.

More nuanced analyses reveal, however, to what extent thinkers such as Edmund Husserl, Martin Heidegger, and Rudolf Carnap were heavily influenced by their Neo-Kantian colleagues. And yet deeper reflections quickly show how much the Neo-Kantian movement as a whole has shaped 20th-century philosophy and beyond. There is currently a revival of Neo-Kantianism, both in philosophical historiography as well as in systematic fields such as epistemology, ethics, and social and political philosophy, which corrects this skewed image. Especially, the work of Michael Friedman on the philosophy of science and its history has been a great force in reviving a central Neo-Kantian project: a priori cognition. Friedman, like no other contemporary philosopher, has enforced the need for a renewed form of transcendental philosophy to meet the challenges of contemporary science. In this intent, he has fruitfully taken up the original concept of the "a priori" developed in the Marburg school for his conception of the "dynamic *a priori.*" Finally, due to the contemporary dearth of historical knowledge of Neo-Kantianism, it is astonishing how many ideas that are currently floated are simply reinventions of the Neo-Kantians' work, such as the topic of normativity, up to current interpretations of Kant's philosophy, which was perhaps the most visible trademark of the Neo-Kantians.

Main Tenets

Despite the heterogeneity of Neo-Kantianism, one can venture some general definitions and indicate some main philosophical paradigms guiding the philosophers who defined themselves as working under the banner of "Kantianism." A general definition of Neo-Kantianism, which is equally utilized by any philosopher today who calls himself a "Neo-Thomist," "Neo-Aristotelian," or the like, is to philosophize in the spirit of Kant but in an updated version of the revered person. This implies rejecting some elements of the orthodoxy in order to salvage certain other aspects that are deemed central and timeless in Kant's work. To be sure, disagreement about what exactly is central in Kant is part of such an adherence to him.

Nevertheless, a more specific definition may be attempted: The Neo-Kantians intended to update the Kantian critical project in light of newer findings, first and foremost in the natural sciences, where the Kantian project is understood as a justification of the *factum* of reason, which is now construed as the undeniable *factum* of the sciences. Furthermore, the main paradigm under which they operated, in the Kantian spirit, can be called *critical* and *idealistic:* critical insofar as the critique is a justificatory enterprise with respect to existing findings (the *facta*) in different spheres of knowledge and idealistic insofar as any realistic scenario in the experience and cognition of anything worldly is rejected. Instead, nearly all problems the Neo-Kantians deal with operate under the basic assumption that there is no direct access to the world as a thing in itself but that all knowledge about the world is mediated by our experience of the world and the conceptual capacities that we have of it. This stance concerns the main topics of inquiry, be they science, cultural artifacts and contexts, or values.

Since it is impossible to give a comprehensive overview of Neo-Kantianism in the space provided, it is best to present this movement in its two main schools in Germany and their general tendencies, respectively. This account will have to pass over the French scene with its main representatives, Émile Boutroux, Léon Brunschvicg, and Jules Lachelier. Suffice it to say that they, too, took the Kantian inspiration to move beyond a materialistic or positivistic worldview, though the French movement was remarkably different due to the tendency on the part of some of its representatives to move in the direction of a spiritualism that would have been quite implausible to its adherents on the eastern side of the Rhine.

The Marburg School

Undoubtedly, the most influential and compact school to emerge was the school centered at the University of Marburg, appropriately called the Marburg school. It included, essentially, Hermann Cohen, Paul Natorp, and Ernst Cassirer, though the latter is sometimes believed to have left the school confines with his original philosophy of symbolic formation. Their main achievement was to work out what they called, in vague adherence to Kant, the transcendental method. This method, first employed by Cohen with respect to the Newtonian sciences, with the added thesis that it was they that Kant had

in mind when he drafted his critical system, intends to reconstruct the cognitive activities of the scientist as she constructs her theories. This is not psychology, however, quite to the contrary; the transcendental philosopher extracts the logical laws that went into the construction of the respective theory; it is a method of "purification" of the work of the scientist to reduce the latter to its bare logical and conceptual elements. If, as the Marburgers claim, "true" reality is found in the theories about nature, transcendental philosophy is equally constructive in grounding the applied theories in a space of pure thought—that is, purely logical structures and a web of logically interconnected concepts. This nexus of logic expands, however, as science progresses, causing Cohen to construe transcendental philosophy as a discipline of grounding cognition in the form of an expanding nexus of categories, which evolve as our knowledge of the world evolves. This way, the Marburgers wanted to preserve Kant's project of a synthetic a priori while acknowledging a progress in the sciences that demands ever new conceptual groundings. The *factum,* in this case of the sciences, thus, was reinterpreted as a *fieri,* a constructing on the part of reason. As mentioned, Friedman has taken up this trope in what he calls "dynamic *a priori.*"

Philosophy, then, has no special space carved out for its own, but it works in close conjunction with the science it concretely justifies. What was exemplified by Cohen, however, in the realm of science was also carried out in other *facta* that make up our culture, namely (following the Kantian architectonics), ethics, art, and religion. Although Cassirer relaxed Cohen's rigidly logical look at reality, he, in keeping with the spirit of Marburg, characterized this project as a move from Kant's critique of reason to a critique of culture. Natorp expanded Cohen's project to also include psychology as a science in the critical purview of this method. In Cassirer, the critique of culture pertained to the plural of "logics" that comprise the different forms in which the human being, as "animal symbolicum," encounters the world as a universe of symbolic formation.

The Baden School

Much less unified was the so-called Southwestern or Baden school, which included Wilhelm Windelband, Heinrich Rickert, and Emil Lask. Their main efforts may be seen as focused on two related issues: (1) a comprehensive theory of science, which included both the human and the natural sciences, and (2) a theory of values, where Plato was just as influential as Kant. As regards the former, especially Windelband and Rickert were focused on the object of cognition, which was to be determined not ontologically but methodologically by focusing on the object either in its lawful generality (nomothetically) or in its individual singularity (idiographically). Depending on the direction of research, the object of cognition is located somewhere between these two ideal poles.

With respect to the much discussed status of values, it was especially Rickert who opted for the values' own sphere of existence, which he called (following Hermann Lotze's interpretation of the Platonic forms) "validity." Perhaps the most promising Southwest Neo-Kantian, Emil Lask died prematurely in the trenches of the Great War but was on the way to working out a category system for philosophy itself (a philosophy of philosophy), which, if completed, could have been hugely important for philosophy in the 20th century.

Concluding Remarks

In this overview are mentioned only the "idealistic" representatives of Neo-Kantianism. This means that there were also some (like Alois Riehl) who favored a realistic reading of Kant's critical philosophy. But given the relative outsider status of Riehl, one can conclude that the main target of the Neo-Kantians was a materialistic worldview as proposed by the experimental sciences, which was seen as reductionistic and ultimately relativistic. They were in favor of an idealistic one, while at the same time remaining close to the actual work in the sciences, critiquing them. In both senses, as making a sustained case for an idealistic and critical philosophy, the Neo-Kantians may be seen as heirs of their namesake.

Sebastian Luft

See also A Priori and A Posteriori; Given, Myth of the; Hegelianism and Contemporary Epistemology; Hermeneutics, Phenomenology, and Meaning; Idealism; *Naturwissenschaften* Versus *Geisteswissenschaften*

Further Readings

Cassirer, E. (2005). Hermann Cohen and the renewal of Kantian philosophy. *Angelaki, 10,* 95–108.

Freidman, M. (2000). *A parting of the ways: Carnap, Cassirer, and Heidegger.* Chicago, IL: Open Court.

Friedman, M. (2001). *Dynamics of reason.* Stanford, CA: CSLI.

Luft, S. (Ed.). (2012). *The Neo-Kantian reader: An anthology of key texts from classical Neo-Kantianism.* London, England: Routledge.

Makkreel, R. A., & Luft, S. (Eds.). (2010). *Neo-Kantianism in contemporary philosophy.* Bloomington: Indiana University Press.

Munk, R. (Ed.). (2005). *Hermann Cohen's critical idealism.* Dordrecht, Netherlands: Springer.

Richardson, A. (2006). "The fact of science" and critique of knowledge: Exact science as problem and resource in Marburg Neo-Kantianism. In M. Friedman & A. Nordmann (Eds.), *The Kantian legacy in nineteenth century science* (pp. 212–226). Cambridge: MIT Press.

Rockmore, T. (2000). *Heidegger, German idealism, and Neo-Kantianism.* Amherst, NY: Humanity Books.

Neo-Marxism

Neo-Marxism is a term that has been used since as far back as 1918. Here, it is taken to mean Marxism as it has survived and developed since the fall of the Soviet Union. This entry discusses the background situation and the main theoretical areas currently discussed.

Marxism was widely thought to have been consigned to the dustbin of history by events around 1989. These included the fall of the Soviet Union, the ending of communism in Eastern Europe, the breakup of Yugoslavia, and the ongoing spread of capitalism in China. In parallel with these developments, communist parties in Western Europe tended to convert themselves into left-wing social democratic parties that accept "bourgeois democracy," meaning that any advances toward communism are susceptible to being reversed by subsequent elections. Intellectual fashion likewise veered away from Marxism, the postwar heyday of which came in the late 1960s and early 1970s, following the events of May 1968 in which France came close to experiencing a socialist revolution. The dominant trend was now postmodernism, with its doctrine that grand theories such as Marxism had died.

However, although initially Marxists were relatively isolated, there has been a growing stream of Marxist theory. This has been particularly spurred on by the virtually worldwide spread of capitalism, which has recently taken a purer form than the Keynesian welfare states that arose following World War II. In the advanced countries, there has been a serious growth of inequality, and in the Third World, the imposition of the so-called Washington Consensus has exacerbated the condition of the poor and widened the gap between rich and poor nations. The fall of the Soviet Union and the intellectual isolation of Marxism has meant that splits between Trotskyists, Stalinists, and Maoists have tended to die away. Enthusiasm for a violent "big bang" revolution has also waned: Extensive bloodshed in order to achieve a reversion to communism in due course does not seem a price worth paying.

A wide range of topics are discussed, as can be seen from the pages of journals such as *Historical Materialism,* an outstanding new initiative in creative Marxist thinking, and *New Left Review.* A particularly widely discussed theme is *justice.* There is much debate as to whether or not Marx had an implicit theory of justice despite his frequent avowals that justice is not a conception he accepts. There is more of a consensus that "scientific socialism" does not actually work very well in various respects. Notably, aspects of Marx's *economics* such as the transformation from value into price and the declining rate of profit are generally held not to work. Linked to this, and to what has actually happened since Marx's day, there is scepticism about the inevitability or even likelihood of revolution, particularly in advanced capitalist states. While it is accepted that an important cleavage in society is that between capital and labor, there is skepticism about the unity of the working class, when those on the minimum wage in the United States earn more than 50 times as much as workers in Third World sweatshops. Much of the workforce in the advanced capitalist countries is, of course, paid well above the minimum wage. Moreover, it can be argued that workers are not as poor as, for example, most carers.

For all these reasons and more, many theorists argue that even if Marx did not have a theory of justice, those working in the Marxist tradition today need one. The question then is what would a Marxist theory of justice look like? It is difficult to answer this. Marx himself says that in a communist society the slogan "From each according to his abilities, to each according to his needs" applies. However, if "needs" is interpreted to mean basic requirements of food, shelter, and other essentials, these are by and large already met in advanced capitalist societies. If it means "anything you want to do," there

are several problems. It is doubtful that the planet's resources will sustain an affluent American lifestyle for everyone. Activities such as making epic films, altering the climate of Australia, or drinking vintage champagne are manageable for some people but not for everyone. Some needs, such as the need for personal care, are indefinitely large.

One of the stronger points of Marxism is its account of the tendency of capitalism to spread itself worldwide, as this is exactly what capitalism is tending to do today. There is then some debate as to the implications of this tendency, notably whether some kind of theory of imperialism is appropriate for them. One much debated theory developed from Marxism juxtaposes the multitude of those who work to the ubiquitous empire of capitalism. However, as seen above, the workforce is too diverse for this to really be plausible.

Not surprisingly, there is debate about issues that have arisen since Marx's day. One of these concerns human nature: Apart from asking whether the young Marx's theory of human nature can be rendered compatible with differences of gender and ethnicity, some theorists ask whether it is valid for cyborgs. The role of the Internet and of an economy based increasingly on the processing of knowledge rather than on straightforward material production is also discussed. More immediate questions such as what is really involved in the war on terror or the invasion of Iraq are of course debated.

Debate continues on a range of issues that were already under discussion back in the 1960s, but with additional aspects that have arisen since then. These include the question of the extent to which various interpretations of historical materialism are compatible with the way societies actually function; the question of whether a Marxist theory of class can be made to fit with social groupings today; the extent to which anything can be salvaged from Marx's economics; the question of whether Marxism works better with its Hegelian heritage emphasized or expunged; related to this, the issue of whether or not Marx's theories really benefit from some version of a dialectic; and the relative quiescence of working people, which in turn points toward the role of ideology and the idea of a labor aristocracy.

Mark Cowling

See also Alienation: From Philosophy to Sociology; Analytical Marxism; Capitalism; Dialectic, in the Social Sciences; Ideology; Marxism and Social/Historical Explanation; Marxist Economics; Marxist Ethics; Reification

Further Readings

Cowling, M. (Ed.). (1998). *The communist manifesto: New interpretations*. New York, NY: New York University Press.

Cowling, M., & Reynolds, P. (Eds.). (2000). *Marxism: The millennium and beyond*. London, England: Palgrave Macmillan.

Cowling, M., & Wilde, L. (Eds.). (1989). *Approaches to Marx*. Milton Keynes, England: Open University Press.

Therborn, G. (2010). *From Marxism to post-Marxism?* London, England: Verso.

Neural Hermeneutics

The term *hermeneutics* originally referred to the art of interpreting complex written texts, in particular holy scriptures, which demand considerable skill to reveal their meaning. Hermeneutics is especially relevant to the problem of translation, which, of necessity, requires interpretation of the original text. The major problem for hermeneutics concerns how to develop criteria for deciding when an interpretation is correct.

Modern practitioners of hermeneutics recognized that this problem applies, more generally, to any situation in which messages have to be interpreted. The philosopher and theologian Friedrich Schleiermacher (1768–1834), among others, suggested that the problem for the interpreter is to reveal the message intended by the speaker—that is, what the author had in mind when she wrote the text (*mens auctoris*). However, this criterion, that the interpretation matches the intention, is problematic since the author might be dead or otherwise unavailable. So all we typically have is the text and some knowledge of the context in which it was written.

This difficulty is not confined to the interpretation of ancient texts. Even if I am talking with you face-to-face, I cannot access your mind to check whether my interpretation of what you have just said corresponds to what you intended me to understand. I can create a coherent story, but I can never get independent evidence about the correctness of my interpretations. Nevertheless, in spite of this apparently insurmountable difficulty, most of

the time people seem to be able to understand each other very adequately. How is this achieved?

Neural hermeneutics is concerned with the mechanisms, instantiated in the brain, through which people are able to understand one another. Such mechanisms, although they might now be specialized for understanding, will have evolved from earlier mechanisms with other purposes. Two such extant mechanisms seem relevant to the problem of understanding. The first is *predictive coding* (or Bayesian inference), which explains our perception of the physical world. The second is that of *simulation* and *alignment,* which aids our perception of the social world.

We perceive an object, such as a tree, on the basis of signals from our senses. This process is not linear, since no sensation can unambiguously indicate the presence of a tree. Rather, a computational loop is required, circling from sensation to inference and back again. Our brain infers the most likely cause of the sensations and then tests this inference by collecting more sensory evidence (e.g., by moving the eyes or touching the object). If the evidence is not what was expected on the basis of the inferred cause of the sensations (a prediction error), then the inference has to be updated. Only when the fit between sensations and inferred cause is sufficiently good is the object unambiguously perceived.

In principle, the same mechanism can be applied when trying to understand the mental world of others. The major difference is that, unlike with trees, the process goes in both directions: While I am trying to understand you, you are trying to understand me. Here, the sensory evidence might be the words I hear, from which I infer the idea you are trying to convey. I can test my inference not only by predicting what else you are likely to say but also by saying something myself and predicting how you will respond. Meanwhile, you will be applying the same strategy to what I say. When our prediction errors become sufficiently low, then we have probably understood one another. In this account, the error we are minimizing is not the difference between my idea and your idea, since we have no direct access to each other's ideas. Rather, it is the difference between my idea and my representation of your idea (see Figure 1).

One advantage of the formulation in terms of predictive coding is that it elegantly captures the concept of the *hermeneutic circle,* originally developed in classic hermeneutics, whereby the whole cannot be understood without reference to the parts, while, at the same time, the parts cannot be understood without reference to the whole. In the same way, in the predictive coding loop, the inferred cause (the idea, *the whole*) predicts the evidence, while, at the same time, the evidence (the words, *the parts*) modifies the inferred cause.

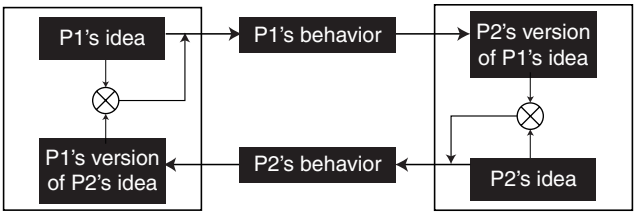


Figure 1 | If there is a discrepancy between P1's idea and P1's version of P2's idea, then P1 will modify her behavior. The two people reach an understanding when they both minimize the difference between their idea and their version of the other's idea.

Source: Chris Frith.

The predictive coding model outlined here does not explain how the link is made between the words and the idea, that is, how the initial inference is made. One possibility is to use *simulation;* that is, I can predict what words you will use on the basis of what I myself would say in the same situation. This is by analogy with motor simulation, in which I predict the movements of others on the basis of my own motor system, a mechanism for which there is now considerable evidence. A necessary consequence of the application of simulation to understanding others is that understanding will be more difficult to achieve if you are in some way different from the person you are trying to understand. This problem may be mitigated through *alignment* (or *mirroring*).

We all have a strong and automatic tendency to imitate each other: the *chameleon effect*. This imitation or mirroring occurs in many domains, including gestures, emotions, and aspects of speech, such as intonation, grammar, and vocabulary. Such mirroring makes us more similar to the person we are interacting with and thereby makes motor and mental simulation more efficient. Direct evidence that understanding is improved by alignment comes from a study showing that communication improved when participants deliberately imitated the accent of the person they were talking to.

The interactive mechanism described above implies that understanding is, in part, a collaboration between the partners engaged in the discourse. Thus, I learn more about my own ideas through interacting with someone else. This relates to Schleiermacher's suggestion that by taking the context into account, the translator can achieve a better understanding of the text even than the original author. By the same argument, a listener can have a better understanding of the speaker than the speaker herself. This is because the listener will not only understand the message that the speaker intends to convey but can also take account of signs, such as body language, indicating aspects of the message that the speaker was unaware of. This better understanding will be fed back to the speaker in the course of the conversation. Thus, through interactions with others, we can achieve a better understanding of ourselves.

Chris Frith and Thomas Schwarz Wentzer

See also Cognitive Phenomenology; Cognitive Sciences; Empathy; Hermeneutics, Phenomenology, and Meaning; Mirror Neurons and Motor Cognition in Action Explanation; Simulation Theory; Social Perception

Further Readings

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science, 21*(12), 1903–1909.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893–910.

Érdi, P., Ujfalussy, B, & Diwadkar, V. (2008). The schizophrenic brain: A broken hermeneutic circle. In V. K rková et al. (Eds.), *Artificial neural networks: ICANN 2008* (pp. 929–938). Berlin, Germany: Springer.

Frith, C. D. (2002). How can we share experiences? Comment from Chris Frith. *Trends in Cognitive Science, 6*(9), 374.

Frith, C. D. (2009). *Making up the mind: How the brain creates our mental world.* New York, NY: Wiley.

Frith, C. D. (2010). What is consciousness for? *Pragmatics & Cognition, 18*(3), 497–551.

Gallagher S. (2004). Hermeneutics and the cognitive sciences. *Journal of Consciousness Studies, 10–11,* 162–174.

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science, 1*(2), 292–304.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27,* 169–192.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin, 131*(3), 460–473.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Science, 10*(7), 301–308.

Neurath's Unity of Science and the Encyclopedia Project

This entry gives an overview of Otto Neurath's thought and, in particular, his vision of a specific variety of unity of science and his famous project of an encyclopedia. The entry first presents Neurath's project, also indicating his political vision behind it; goes on to unravel the general doctrine of the unity of science in its various forms; and ends by explaining Neurath's own distinctive conception of scientific unity. Neurath's ideas were specifically concerned with the social sciences and their relation to the natural sciences.

Historical Background

The doctrine of *the unity of science* was the mainstay of the philosophical movement known as logical positivism and/or logical empiricism. According to this doctrine—and contrary to Neo-Kantianism, Phenomenology, and Neo-Wittgensteinianism—there is no sharp, categorical distinction to be drawn between the natural sciences on the one side and social science or the human sciences on the other. Often understood to be reductive in its intent and caricatured as aiming to replace social science by a future physics, the doctrine took a particularly interesting and decidedly antireductivist form in the hands of Otto Neurath, who anticipated in some respects the criticisms of contemporary opponents of the unity of science doctrine.

Neurath's Project

Otto Neurath (1882–1945) was a member of the Vienna Circle group of philosophers that met between 1924 and 1936 (other members included Moritz Schlick, Rudolf Carnap, Philipp Frank, and

Hans Hahn) and that became famous for its opposition to metaphysics as cognitively meaningless (they conceded that it served other psychic needs). In this group, Neurath consistently opposed phenomenalist epistemologies and stressed the ever hypothetical and holistic character of our knowledge. For him, philosophy became science in a critical self-reflective mode aiming for its systematic development. The goal was the *unification,* and later the "orchestration," of its multiple branches, which, in turn, was to be documented by the multivolume project of the *International Encyclopedia of Unified Science* (IEUS). (The 20 monographs that appeared between 1938 and 1970 have since been collected as the two-volume *Foundations of the Unity of Science.*)

A social scientist by training and an ardent social reformer by conviction, Neurath wanted social science to contribute to the *reshaping of society.* Concerned with its applicability, he made *empirical testability* the hallmark of its theories. What he primarily opposed as metaphysics was high-flown obfuscation serving to mask social inequality and exploitation. The unity that he aimed for has aptly been characterized by Nancy Cartwright as "unity at the point of action." In this, as in his understanding of other doctrines characteristic of the movement, Neurath's views were increasingly at odds with the emerging orthodoxy of logical empiricism. The latter's strictly formalist orientation toward philosophy of science cared little for the complementation of the study of the logic of science (the logic of scientific representation) with the study of the pragmatics of science (its conditions and exigencies as a social activity), which Neurath thought essential.

The Rise and Fall of the Unity of Science Doctrine

Different versions of the doctrine of unified science were affirmed by different protagonists at different times, reflecting the extremely rapid development of the theories proposed by the logical empiricist movement in its first two decades—before, that is, a certain stasis set in after the end of World War II, with its reputed establishment as the dominant form of Anglo-American philosophy of science. It is notable that one of the most reductivist forms of the doctrine was propounded by the contemporary American philosopher Hilary Putnam in 1958 (a stepwise reduction of theories all the way down to physics), at a time when Carnap had already abandoned the ideal of a reductive hierarchy of laws of different scientific disciplines. Yet even what Carnap retained—the unity of method and the unity of language—was still too ambitious for the historicist opposition to logical empiricism that gained ground in the aftermath of the publication of Kuhn's *The Structure of Scientific Revolutions* in 1962. (It should be stressed that Kuhn's work was published in precisely the IEUS that Neurath had started.) What is clear is that by that time Neurath's alternative conception of the unity of science had been lost from view. What remains unclear is whether, had it been known, it would have made a difference, given the antipositivist fervor that has since held sway.

Then, as now, opposition to the unity of science thesis tended to be motivated at least in part by concern for retaining the distinctly human dimension of the nonnatural sciences. Yet this concern had also already strongly motivated the separatist doctrines that were propounded with particular intensity in the German-speaking parts of Europe when the logical empiricists opposed them (as then also supporting reactionary political ideologies). This concern focused attention on the diversity of the types of objects of the natural and the social sciences (nonminded vs. minded, i.e., physical events vs. human beings), on the diversity of their epistemic aims (nomological explanation vs. understanding of singular cases), or on the diversity of the methods of investigation they used (experimental testing vs. hermeneutical interpretation)—or on a combination of these—stressing that one account of science would not fit all. Given the difficulty of denying any differences between the natural and the social sciences on these accounts, the question arises how a unity of science thesis can accommodate them.

Neurath's Conception of the Unity of Science

Neurath's answer emphasized the aspects of the unity of the method of science and the unity of the language of science but discounted that of the unity of its laws. Instead of aiming for the successive reduction of the laws of the special sciences to those of fundamental physics, Neurath stressed that all that was needed were cross-connections between the sciences, allowing for the explanation

(and, where possible, prediction) of individual or types of events by the combination of the resources of different sciences. (He was not disturbed by the limits of predictability in the social domain and later even stressed that the occasional unpredictability in principle held across all sciences.) In addition, he pointed out that the generalizations of the social sciences often failed to amount to the universal ones familiar from physics, holding instead only for particular historical periods and cultures or types of societies. Importantly, unlike his more orthodox colleagues and opponents, like Karl Popper, he defended this limitation as legitimate and opposed the tendency to use the universal laws of physics as the deciding measure. Most significantly, Neurath opposed the idea of unified science as a deductively closed system of axiomatized theories as an impractical and misleading ideal and instead championed the model of an "encyclopedic," case-by-case integration of the sciences where possible. (Needless to say, not all contributions to the IEUS shared this vision.)

The *autonomy of the social sciences* was further secured by Neurath's conception of the unity of the language of science. For him, this unity did not consist in the eliminative reducibility of all scientific predicates to those of mathematical physics or even to those of the "intersubjective thing-language" but in the fact that all scientific statements had to be checked for their logical compatibility with so-called protocol sentences. Unlike Carnap's bare observation statements, Neurath's protocol sentences were canonically regimented statements of scientific testimonies, indicating sets of conditions for the acceptance of the testimonial claims made. All that was therefore required of the language of legitimate social sciences was that their claims could be checked in this way. Of particular importance here was that Neurath, unlike Carnap, also allowed these protocols to contain predicates that were intersubjectively confirmable (by indirect means) but not intersubjectively testable (by direct means). This allowed protocols to feature reports of psychological states and thus allowed for social-scientific theories that dealt in the beliefs, preferences, plans, and intentions of individual agents to be checked in practice—without, that is, reliance on questionable logically behaviorist definitions or unknown neurophysiological identity conditions for the referents of those intentional locutions.

In consequence, Neurath's conception of the unity of method was very liberal. When he rejected the empathy-based conceptions of a separatist *Geisteswissenschaft,* he rejected the claim that empathetic understanding was necessary in general and ever sufficient for social-scientific knowledge, but he did not deny that empathy could be used as a heuristic of discovery as long as it was recognized that its findings required independent justification. Also, interpretation in a more general sense (attribution of intentional states to agents) was a legitimate social-scientific method by Neurath's lights. What he opposed as Max Weber's metaphysical leanings was not Weber's interpretive sociology but his attempt to reverse the order of explanation postulated by historical materialism. That is, Neurath felt that Weber's insistence on meaningful behavior as an explanation of human action in the social world was alright up to the point that it did not undermine the robust material causal mechanism postulated by Marxism.

The unity of scientific method that the social sciences had to hew to thus consisted in a variation of Carnap's conventionalist take on the principle of empiricism: to formulate the findings of social science in such a way that they are intersubjectively confirmable by means of scientific testimonies.

Conclusion

Whether Neurath's own conception of the unity of science can overcome or undermine all the qualms of contemporary theorists of the disunity of science cannot be decided here. It may be noted, however, that the latter's objections to the traditional conception of the unity of science mostly bypass Neurath's version.

Thomas Uebel

See also Hypothetico-Deductivism; Logical Positivism/Logical Empiricism; *Naturwissenschaften* Versus *Geisteswissenschaften;* Neo-Kantianism; Reduction and the Unity of Science; Sociology of Knowledge and Science; Verificationism

Further Readings

Carnap, R., Morris, C., & Neurath, O. (Eds.). (1970). *Foundations of the unity of science.* Chicago, IL: University of Chicago Press.

Cartwright, N., Cat, J., Fleck, L., & Uebel, T. (1996). *Otto Neurath: Philosophy between science and politics*. Cambridge, England: Cambridge University Press.

Galison, P., & Stump, D. (Eds.). (1996). *The disunity of science*. Stanford, CA: Stanford University Press.

Neurath, O. (1983). *Philosophical papers* (R. S. Cohen & M. Neurath, Eds.). Dordrecht, Netherlands: Reidel.

Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science: Vol. 2. Concepts, theories, and the mind-body problem* (pp. 3–37). Minneapolis: University of Minnesota Press.

Reisch, G. (2005). *How the Cold War transformed philosophy of science*. Cambridge, England: Cambridge University Press.

Symons, J., Pombo, O., & Torres, M. (Eds.). (2011). *Otto Neurath and the unity of science*. Dordrecht, Netherlands: Springer.

Uebel, T. (2007). Philosophy of social science in early logical empiricism: The case of radical physicalism. In A. Richardson & T. Uebel (Eds.), *The Cambridge companion to logical empiricism* (pp. 250–277). Cambridge, England: Cambridge University Press.

Neuroeconomics

Neuroeconomics is an emerging interdisciplinary field combining economics and neuroscience. Neuroeconomics investigates various problems, combining economics with diverse neuroscientific studies. In particular, neuroimaging technologies such as positron emission tomography and functional magnetic resonance imaging have played an important role in the development of neuroeconomics, although other studies, such as lesion studies and physiology, are also indispensable for neuroeconomics. These neuroimaging technologies enable us to noninvasively investigate how human brains work when human beings are asked to do some task. This entry presents the novel field of neuroeconomics and reviews the two dominant approaches in it. Both approaches raise a number of epistemological issues.

Some might think that neuroeconomics is a unified field of economics and neuroscience, but economists and neuroscientists take different approaches to neuroeconomics. The first approach performs behavioral economic experiments using neurologic imaging studies to criticize standard economics. The second approach uses standard economics to develop algorithmic models of neural structures. This difference of attitude toward standard economics reflects the fact that economists and neuroscientists have different interests in neuroeconomics, and thus this could potentially affect the future of neuroeconomics as a field.

The Behavioral Economic Approach to Neuroeconomics

Some prominent behavioral economists, such as Colin Camerer and Ernst Fehr, take this approach to neuroeconomics. Their findings have generated much of the current interest in neuroeconomics. They employ game-theoretic situations, such as prisoner's dilemma games, ultimatum games, and trust games, and investigate how subjects' brains work. For instance, Dominique de Quervain and his colleagues in Fehr's research group have investigated whether there is any neural basis for altruistic punishment of defectors in the trust game. According to this experiment, when subjects are asked whether they would spend money to punish defectors who betray trust, the subjects' dorsal striatum, an integral part of the brain's reward system, is activated. Furthermore, subjects with greater activation of the dorsal striatum spend more money for punishment. This suggests that the punisher derives satisfaction from punishing the defector who violates social norms. Such a finding is expected to reveal the neural underpinnings of human cooperative behavior.

Neuroeconomists in the first approach share a motivation with behavioral economists. They intend to reveal how real human behaviors differ from theoretical assumptions and, thus, to criticize the idea of homo economicus, or economic man. Although it is well-known that the idea of homo economicus differs from flesh-and-blood human beings, it has been used as the main assumption in standard economics for many years. By criticizing the idea of homo economicus, neuroeconomists using this approach aim to modify standard economics. But scholars such as Herbert Simon have already pointed out the discrepancy between theory and reality and proposed some alternatives to overcome it, so the idea of neuroeconomics is not entirely new. Armed with the latest developments in neuroimaging technologies, neuroeconomists hope to provide a better alternative to standard economics.

The Neuroscientific Approach to Neuroeconomics

According to the second approach, neural tissues *economically* evaluate what reward would be obtained from a behavior, based on information such as internal states and/or external stimuli. This process is necessary for our decision making because our resources are limited. To see neural tissues in such a way is to *regard neural activities as an object of economic analysis*. This is why some neuroscientists become interested in economics, and thus they are led to neuroeconomics. But those who are interested in this second approach do not necessarily intend to criticize standard economics. Rather, researchers such as Paul Glimcher positively use standard economics in their studies of neural activities. Among others, they try to develop a mathematical model of the brain's reward system, in particular the role of dopaminergic neurons in predictive valuation and learning. Based on such a study, they aim at constructing a comprehensive algorithmic model of the brain. True, such a model may be just a model, because neural activities do not necessarily follow theoretical assumptions. But contrary to neuroeconomists in the behavioral economics vein, neuroeconomists in the second approach do not find this problematic. In their view, the mathematical model helps understand how reality differs from theory. In this sense, we could say that neuroeconomists using the second approach regard the mathematical model as an ideal type. Such a stance toward standard economics differs from the behavioral economic approach.

Quo Vadis Neuroeconomics?

As described above, there are two different approaches to neuroeconomics. Although major neuroeconomists know that this is a problem, they hope to arrive at some convergence. But the future of neuroeconomics is not clear for reasons other than the reason suggested above. One is that the results of noninvasive neuroimaging technologies on which many neuroeconomic studies rely are correlational. To do a causal explanation, neuroeconomists need to use noninvasive brain stimulation technologies, such as transcranial magnetic stimulation, which can activate or inactivate particular parts of the brain. But because of the risk of possible damage to the brain, Leslie Sargent Jones warns against the use of brain stimulation technologies on human subjects. This would affect the behavioral economic approach, because it is mainly interested in *human* behaviors in economic situations. The neuroscientific approach, however, performs experiments not only on human subjects but also on animals. Thus, the latter may be more productive than the former. But given the European Union's recent legislations on animal experiments, this might be too optimistic. We are yet to see how this new field develops and whether there can be any convergence between two different approaches.

Kei Yoshida

See also Biology and the Social Sciences; Experiments in the Social Sciences: Ethical Issues; Game-Theoretic Modeling; Heterodox Economics; Homo Economicus; Mathematical Models, Use in the Social Sciences; Naturalism in Social Science; Neuroethics; Neuroscience and Politics; Social Neuroscience

Further Readings

de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science, 305,* 1254–1258.

Glimcher, P. W. (2003). *Decisions, uncertainty, and the brain: The science of neuroeconomics.* Cambridge: MIT Press.

Glimcher, P. W., Camerer, C. F., Fehr, E., & Poldrack, R. A. (Eds.). (2009). *Neuroeconomics: Decision making and the brain.* Amsterdam, Netherlands: Elsevier.

Jones, L. S. (2007). The ethics of transcranial magnetic stimulation. *Science, 315,* 1663.

Ross, D. (2008). Two styles of neuroeconomics. *Economics and Philosophy, 24,* 473–483.

Neuroethics

Neuroethics is a newly emerging branch of philosophy, with two main focuses: (1) ethical issues arising out of the sciences of the mind (especially, but not only, neuroscience) and (2) the ways in which these same sciences illuminate traditional philosophical issues such as the existence of free will, the nature of personhood, and the features of morality. Neuroethics is a heavily interdisciplinary area of inquiry, bringing together researchers from the humanities and the natural and social sciences.

This entry presents these two areas of neuroethics, explains the role of technology, highlights new findings from experimental social science challenging the received view that moral judgments are rationally arrived at (based on a supposedly universal rational intuition all humans have), shows how this approach brings neuroethics close to so-called experimental philosophy, and, what is more important, underlines the increasingly dominant role of various branches of social science in transforming how philosophy looks at moral judgments (the second area of neuroethical concern).

Beyond Bioethical Issues

Much of the subject matter and approach of neuroethics is closely akin to the subject matter and approach of bioethics. Whereas bioethics is concerned with ethical issues arising out of the biomedical sciences and their application, ranging from the permissibility of abortion to whether life support may ethically be withdrawn, much of neuroethics concerns the applications of our growing knowledge of the brain. Neuroethicists consider questions such as the permissibility of the use of cognitive enhancements, the potential loss of privacy from the development of neuroimaging-based systems of "brain reading," and the manipulation of memory using psychopharmaceuticals. These questions are urgent and intrinsically interesting, but they are not in themselves of particular interest to social scientists.

Technology

However, an adequate analysis of the ethical issues with which neuroethics deals requires a detailed knowledge of the powers and limits of the technologies whose application is under consideration, and that requires that these technologies be understood. The technologies are typically developed by researchers in the natural sciences, such as neuroscience and biochemistry, but the theories these researchers draw upon stem largely from psychology. Furthermore, assessing the permissibility of neuroethical technologies often requires expertise in other social sciences, such as sociology and economics. For instance, assessing the common claim that neuroethics raises issues of social justice requires that we be able to gauge the likelihood that cognitive enhancements will be available only or mainly to the already better-off individuals or whether they are likely instead to rapidly become cheap enough to be available also to individuals of lower socioeconomic status. An understanding of the mechanisms whereby other technologies, such as cell phones, have diffused across societies may enable us to make better-informed predictions with regard to this question.

Experimenting on Moral Judgments

Expertise in the social sciences, and especially in psychology, is also required for engagement in the other branch of neuroethics, the understanding of the ways in which the sciences of the mind illuminate traditional philosophical concerns. In a number of areas, philosophers have relied upon intuitions to understand the structure of the concepts they deploy; that is, they have consulted their gut reactions or a priori judgments when trying to delineate concepts. This way of proceeding has often been justified by the claim that competent speakers of a natural language have an implicit understanding of the extension of their concepts. However, work in linguistics and in psychology casts doubt on this justification. Many linguists believe that concepts are not structured as sets of implicit necessary and sufficient conditions; rather, conceptual classification is achieved by reference to a prototype (e.g., what it means to be a "foetus"). If that is the case, then there may be no determinate answer to whether a certain entity falls within the extension of a concept (thus altering the way moral judgments concerning its treatment are to be accepted or not). Furthermore, a wide range of work in psychology suggests that many intuitive or a priori judgments are the product of cognitive biases or of heuristics, which, while normally reliable, may misfire in particular cases. Neuroethicists have therefore turned to evidence from the social sciences to explore when our intuitions about ethical issues should be disregarded as unreliable.

Experimental Moral Cognition: Some Recent Challenges

Recent work on moral cognition has proved a fruitful source of inspiration for neuroethicists pursuing this agenda. Consider Jonathan Haidt's work, for instance. Haidt argues that his experimental evidence supports the view that moral judgments are generated by subjects' emotional responses to cases and that the role of reasoned argument is merely to offer support to our judgments. It follows that

moral judgments will be relatively impervious to being undermined by reason. Haidt's evidence for this claim consists in the fact that his subjects do not typically back away from their moral judgments when they cannot offer adequate grounds for them and that judgments can be manipulated by manipulating emotional responses (e.g., by post-hypnotic suggestion). On the assumption that Haidt has demonstrated that moral judgments can be caused by irrelevant features of cases, we might hope to show that particular classes of intuitions should be rejected as unreliable, on the ground that they respond to irrelevant features of cases. Alternatively, we might conclude that the evidence from social psychology shows that we cannot rely on intuitions at all and must, instead, generate moral judgments by applying an algorithm. Some consequentialists have responded to Haidt's claims in that kind of way.

Of course, Haidt's work represents just one source of inspiration for neuroethical work on moral judgment; other sources include the rich vein of work on alleged *moral modules,* older work in the heuristic and biases traditions detailing particular ways in which judgments may go wrong, and work in the *situationist* tradition in psychology. Other work in psychology promises to illuminate other neuroethical concerns. There is a large body of work in social, cognitive, and developmental psychology on *self-control,* which has been thought by some neuroethicists to be essential to a better understanding of weakness of the will; other work bears importantly on the nature of the sense of self, the possibility of self-knowledge, and even the existence of free will.

The Ascending Role of Social Sciences

It should be noted that neither the topics just mentioned nor the conviction that they should be approached with at least one eye on the social sciences are the exclusive concern of thinkers who describe themselves as neuroethicists. Philosophers of mind, moral philosophers, and philosophers of biology, among others, make use of this kind of data for similar ends. Neuroethics fades into what has been called by some thinkers *empirical philosophy* and even into the new movement of "experimental philosophy," which actively generates empirical results for further reflection; insofar as it can be distinguished from these approaches, it is by the degree of concern that workers in neuroethics evince for the moral and social implications of the work they perform.

Neil Levy

See also Evolutionary Ethics; Experimental Philosophy; Free Will, Philosophical Conceptions of; Moral Cognitivism; Neuroeconomics; Neuroscience and Politics; Thought Experiments; Transhumanism and Human Enhancement

Further Readings

Farah, M., & Wolpe, P. R. (2004). New neuroscience technologies and their ethical implications. *Hastings Center Report, 34,* 35–45.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108,* 814–834.

Illes, J. (Ed.). (2005). *Neuroethics: Defining the issues in theory, practice and policy.* Oxford, England: Oxford University Press.

Levy, N. (2007). *Neuroethics: Challenges for the 21st century.* Cambridge, England: Cambridge University Press.

Roskies, A. (2002). Neuroethics for the new millenium. *Neuron, 35,* 21–23.

Sinnott-Armstrong, W. (Ed.). (2008). *Moral psychology: Vol. 2. The cognitive science of morality: Intuition and diversity.* Cambridge: MIT Press.

Neuroscience and Politics

This entry addresses the application of methods and models from cognitive neuroscience to political questions and problems. The majority of these examinations have looked at the differences between political liberals and conservatives or have sought to provide less biased means to study socially sensitive topics such as race. Most of these studies involve the application of functional magnetic resonance imaging (fMRI) to understand more about the human brain, but neuroscientific methods also encompass hormonal, physiological, and genetic explorations.

The following section discusses fMRI technology and provides some examples of work that has used it to investigate political topics. The entry concludes with a brief overview of other neuroscientific methods that have been employed to study political topics.

Functional Magnetic Resonance Imaging

Neuroscientific methods allow the possibility of exploring some aspects of brain functioning that were not possible prior to the advent of more advanced technology. Brain imaging uses differences in the level of blood oxygen to impute regions of relative brain activity. This analysis allows observers to determine with great accuracy the geographical region in which the greatest activity takes place during a given task. However, this technology cannot provide as accurate a temporal record of when this activity takes place as other methods, such as the use of the electroencephalograph, which typically employs a cap on the head to record the speed of neural activity. Morever, while fMRI has provided untold information regarding the anatomy of much brain activity, it cannot alone explicate the pathways and mechanisms involved in any given activity. In fact, many individuals may use different brain regions to solve the same task, just as many others may use the same regions to respond to different challenges. In this way, brain imaging can provide a more direct measure of response. Finally, it is not necessarily the case that such studies provide more useful or valid information than that derived from psychological studies that involve the use of other methods. As with any other methodology, the utility of fMRI technology depends in part on the question being asked and the kind of evidence sought in order to answer it.

However, fMRI technology has been used to examine some important aspects of *political preferences* and *behaviors*. Work in this area has shown, for example, that political conservatives and liberals appear to have different brain structure. Other work has shown that liberals and conservatives differ in their evaluative processes in risk taking as well. Most of this work implicates the *emotional centers* of the brain as central to the processing and interpretation of political information, including the evaluation of political candidates and policy issues.

Much of the work in this area has examined how individuals respond to political candidates. Influential work has demonstrated that negative attributions appear more important than positive ones in mediating the effect of appearance on voters' decisions. While attractiveness typically enhances a candidate's appeal, this characteristic does not appear to be able to overcome assessments of competence or threat.

In addition, fMRI technology has been used to examine how people respond to sensitive topics such as *race,* where individuals may prove either unable or unlikely to tell the truth about how they feel. These studies have shown that people process facial emotions differently based on the race of the viewer and the subjects. One of the most provocative of these studies showed increased activation in the amygdala, a part of the brain understood to be involved in emotional learning, processing, and evaluation, when White subjects looked at pictures of Black faces. However, this effect disappeared when subjects looked at the faces of famous or positively regarded Black figures.

Other Neuroscientific Analysis

Tremendous advances have been made recently in studying the basis of political preferences and behavior using *genetic* methods and models. These studies can use a variety of techniques to examine the nature of individual variance in the outcome of interest. Genetic studies often involve the genotyping and in-depth interviewing of related family members, including twins, nontwin siblings, parents, spouses, and children. Such work can also involve the intensive investigation of the relationship between genetic variance on specific genes or genetic pathways, and particular outcomes of interest. Scholarship in this area has demonstrated, for example, that more than *half the variance in political ideology can be attributed to genetic causes.*

Additional work has examined the relationship between *hormonal* factors and political outcomes. Early work in this area explored the role of serotonin in dominance behavior, while more recent work has examined the influence of testosterone on aggression or of cortisol on voting behavior.

Other methods designed to assess the neurological responses of individuals to political topics include tests such as the implicit association test, which is designed to examine small differences in reaction time between positive and negative associations to key concepts or triggers, such as race. Studies that use reaction time alone have also produced influential results, showing, for example, that 1-second judgments of competence based entirely on facial appearance can predict the outcome of senatorial races better than chance and can directly track the margin of victory in such races.

Physiological measures, using strategies such as rapidity of eye blink in the face of threat, have also been used to examine the differences between political liberals and conservatives. Other physiological

measures include galvanic skin response, which measures small changes in the secretion of sweat on the skin, or eye-tracking technology, which examines the areas that individuals look at in a given image, how long they spend looking at that region, and how often they return to examine it further. Work in these areas has often found systematic differences in the response patterns of political liberals and conservatives.

Rose McDermott

See also Affective Intelligence in the Social Sciences; Biology and the Social Sciences; Cognitive Sciences; Evolutionary Political Science; Genetic Indeterminism of Social Action; Mirror Neurons and Motor Cognition in Action Explanation; Neuroeconomics; Neuroethics; Political Psychology; Social Neuroscience

Further Readings

Hatemi, P. K., Hibbing, J. R., Medland, S. E., Keller, M. C., Alford, J. R., Smith, K. B., . . . Eaves, L. J. (2010). Not by twins alone: Using the extended family design to investigate genetic influence on political beliefs. *American Journal of Political Science, 54,* 798–814.

Mattes, K., Spezio, M., Kim, H., Todorov, A., Adolphs, R., & Alvarez, R. M. (2010). Predicting election outcomes from positive and negative trait assessments of candidate images. *Political Psychology, 31,* 41–58.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12*(5), 729–738.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*(5728), 1623–1626.

New Wittgensteinians

"New Wittgensteinian" is a moniker used to pick out the members of a particular set of readers of Wittgenstein. Commentators on Wittgenstein's philosophy generally speak of a dramatic rift between his early and later conceptions of the workings of language. In contrast, some of the philosophers who are referred to as New Wittgensteinians claim that there are fundamental *continuities* between his earlier and later accounts of linguistic phenomena. While, on the one hand, these philosophers emphasize that there are fundamental similarities between Wittgenstein's philosophical preoccupations at different periods, on the other, they hold that there are certain telling differences and that his thought about language receives its most compelling treatment later on. Their claims about Wittgenstein's later view of language, even when considered in isolation from their claims about the evolution of his thought, represent a point of substantial conflict with the more widely received interpretations. Interest in the basic view of language in question is what unites the work of the different philosophers who get dubbed New Wittgensteinians, some of whom have little or nothing to say about Wittgenstein's early work. At issue is a view of language that, if sound, would have significant implications for how we approach central questions of philosophy and the social sciences, including, first and foremost, questions about the logical character of the resources available to us for critically assessing social practices.

This entry briefly describes the relevant view of language, distinguishes it from superficially similar views sometimes attributed to the later Wittgenstein, and discusses its bearing on these questions.

New Wittgensteinians: Realism and Anti-Realism

Wittgenstein is frequently read as having espoused a form of philosophical realism at the time of the writing of his early work, the *Tractatus,* and as having exchanged it for a form of anti-realism later on. When depicted as championing philosophical realism, the author of the *Tractatus* is interpreted as favoring a conception of discourse on which the meanings of expressions are determined by ties to objects in a prior reality. When portrayed as trading in a realist outlook for a species of anti-realism, the later Wittgenstein is taken not only to be rejecting the thought that meaning is determined in this way but also to be claiming that there can therefore be no such thing as objective agreement between language and the world. The point is that meaning is fixed not by links to an antecedent reality but by our public practices with expressions and, moreover, that, because meaning is thus a function of what we do, there can be no question of its attaining our ideal of objectivity.

While not all the philosophers who have been labeled New Wittgensteinians discuss the *Tractatus,* those who do are committed to disputing this narrative about Wittgenstein's philosophical development.

They take Wittgenstein to be, in the *Tractatus* as well as in later writings, concerned with repudiating a presupposition *common* to the sorts of realist and anti-realist doctrines that are attributed to him at different periods. Read as a realist, the author of the *Tractatus* is taken to be presupposing that we have the sort of transcendent perspective on language from which to determine that any objective authority it possesses is a matter of grounding in an antecedent reality. And an important claim of the "new" body of work on the *Tractatus* is that its author not only doesn't make this presupposition but sets out to expose it as confused. Furthermore, according to the pertinent New Wittgensteinians, Wittgenstein's early writings anticipate his later philosophical efforts in this respect.

Different anti-realisms are often depicted as anti-metaphysical enterprises, so it might seem as though the idea of hostility to transcendent metaphysics is unthreatening to familiar anti-realist readings of Wittgenstein's later work. But the anti-realist posture ascribed to the later Wittgenstein resembles the realist posture ascribed to him at the time of the *Tractatus* in making the following metaphysical presupposition, namely, that whatever objective authority, if any, language has is a matter of a foothold in a prior reality. Admittedly, the supposedly Wittgensteinian anti-realist differs from her realist counterpart in denying that there are transcendent objects underwriting our modes of thought and speech. But this thinker represents our claim to objective authority as turning on the existence of such objects, and she accordingly presupposes that we can grasp the notion of an external vantage point on language, at least well enough to understand what the—in her eyes unachievable—ideal of objectivity is like. This is the metaphysical presupposition, common to realism and anti-realism, that some New Wittgensteinians take Wittgenstein to be challenging both in his early and in his later writings.

What unites the work of all New Wittgensteinians, those who do as well as those who do not discuss the *Tractatus,* is a sympathetic concern with Wittgenstein's later efforts to arrive at a conception of our lives with language that is not beholden to the presupposition of a transcendent perspective. Wittgenstein is taken to be treating the idea of such a perspective on language as utterly confused, so that now there can be no question of regarding its forfeiture as threatening our entitlement to the ideal of objectivity. By the same token, he is taken to be representing sensitivities characteristic of us as people who have mastered a set of linguistic practices as contributing internally to our ability to think rationally about the world and bring it accurately into focus.

Implications for the Philosophy of Social Sciences

It is insofar as "new" readings of Wittgenstein develop this image of our rational capacities that they have a significant bearing on philosophical debates about our resources for critically assessing the actions, practices, and institutions that are the objects of the *social sciences*. It is an, at least, implicit assumption of many contributions to these debates that we are obliged to accept some version of one of the two following basic positions.

The centerpiece of the first position is the view that there are universal standards for rationality and, further, that these standards are such that our ability to recognize their correctness and apply them does not depend on our having an appreciation of the social practices at issue. This view is attractive insofar as it seems to equip us to offer wholeheartedly rational assessments of social practices. Yet it is unattractive insofar as it also seems to enable us to do so from outside, in what some would impugn as a hopelessly paternalistic manner.

The second basic position that figures in debates about our resources for criticizing social practices pivots on the view that standards for the assessment of given practices are *internal* to the particular practices themselves. This view is attractive insofar as it makes assessment inseparable from a close study of particular practices. Yet it is unattractive insofar as it excludes the possibility of rationally authoritative assessment, thereby ushering in an extreme form of cultural relativism.

What gives the image of our rational capacities characteristic of New Wittgensteinians interest in this context is the fact that it suggests the possibility of combining the virtues of the two foregoing positions while also eliminating their vices. Now it appears possible consistently to say both that grasping standards appropriate to assessing a practice necessarily involves entering into the concerns and interests that shape it and, additionally, that this does not disqualify the standards we thereby pursue from being fully rational. There appears to be a possibility here because, against the backdrop of the image of our rational capacities favored by New Wittgensteinians,

there is no question of a transcendent standpoint from which any interests or concerns we explore have a necessary tendency to obstruct our view of what a practice is really like. The upshot is that New Wittgensteinians' notion of a shared understanding of our rational capacities holds forth the promise of radically reshaping the conceptual space in which debates about rationality within the philosophy of the social sciences take place.

Distinguishing New Wittgensteinians From Older Approaches

These issues have to some extent been obscured by disputes about the legacy of Peter Winch. The disputes in question surround Winch's widely discussed 1958 monograph *The Idea of a Social Science and Its Relation to Philosophy.* In this book, as well as in various later reflections on it, Winch presents himself as drawing his chief inspiration from Wittgenstein's later reflections on language. There is a general consensus among readers that Winch is rightly read as favoring some version of what is described above as an anti-realist reading of Wittgenstein's later philosophy and that Winch's work therefore falls squarely on the relativistic side of familiar debates about our resources for criticizing social practices. This consensus has contributed to an intellectual climate in which a friendly interest in Wittgenstein's later view of language is sometimes regarded as inseparable from a commitment to cultural relativism. The existence of such a climate represents an obstacle to registering the significance of the writings of New Wittgensteinians. But it remains the case that this body of writing has the potential to reconfigure key conversations in the philosophy of the social sciences.

Alice Crary

See also Holism, in the Philosophy of Language; Language-Games and Forms of Life; Rationality and Social Explanation; Realism and Anti-Realism in the Social Sciences; Relativism and the Social Sciences: From the Sapir-Whorf Hypothesis to Peter Winch

Further Readings

Cavell, S. (1999). *The claim of reason: Wittgenstein, skepticism, morality and tragedy.* Oxford, England: Oxford University Press.

Conant, J. (2002). The method of the *Tractatus*. In E. H. Reck (Ed.), *From Frege to Wittgenstein: Perspectives on early analytic philosophy* (pp. 374–462). Oxford, England: Oxford University Press.

Crary, A., & Read, R. (2000). *The new Wittgenstein.* London, England: Routledge.

Diamond, C. (1991). *The realistic spirit: Wittgenstein, philosophy and the mind.* Cambridge: MIT Press.

McDowell, J. (1994). *Mind and world.* Cambridge, MA: Harvard University Press.

Putnam, H. (1999). *The threefold cord: Mind, body and world.* New York, NY: Columbia University Press.

Newtonianism in Adam Smith's Social Science

Isaac Newton's influence on Adam Smith has been widely acknowledged, but its real nature is still a matter of debate. Within the generally accepted view, Adam Smith would have applied Newtonianism to social science, setting the ground for neoclassical and modern mainstream economics. But the reality of this assertion is disputable.

This entry reviews and assesses the thesis that Adam Smith's conception of social science was influenced by, or was a direct descendant of, Newton's scientific method.

Newton and Newtonianism

In 1687, Isaac Newton, the father of modern physics, published *Philosophiae Naturalis Principia Mathematica* (*Principia*). In 1704, a year after the death of his rival Robert Hooke, Newton published in English his *Opticks: or, A Treatise on the Reflexions, Refractions, Inflexions and Colours of Light*. The irrefutable character of *Principia,* with its laws of motion and the universal law of gravity "derived from phenomena," became paradigmatic during the 18th century. And the spectacular nature of Newton's *Opticks,* with its experimental results and its many speculations, influenced the way modern philosophers would think about the world. To assess Newton's monumental influence, we must simply recall Alexander Pope's intended epitaph for Newton: "Nature and nature's laws lay hid in night: / God said, Let Newton be! And all was light" (1730).

Newton's impressive discoveries in natural philosophy influenced a long intellectual tradition. As Newton found the key to understanding nature with his "experimental philosophy," many intellectuals relied on his method to uncover the nature of social

sciences. During the Enlightenment, philosophers would rely on Newton, assuming that social phenomena could be understood using his methodology. In Query 31 of his *Opticks,* the father of modern science declared, "And if natural Philosophy and all its Parts, by pursuing this Method, shall at length be perfected, the Bounds of Moral Philosophy will also be enlarged." This statement was taken seriously in the 18th century.

But the image of Newton as the father of the Age of Reason is biased. The economist John Maynard Keynes, after reading some of Newton's manuscripts, wrote his essay "Newton, the Man," challenging the glorified image of the father of modern science. In the past decades, a renewed interest in the figure, the discoveries, and the methodology of Newton has developed. If *Principia* and the *Opticks* are the most important "public" sources for understanding Newton's method, today we know that Newton also conceived his speculations by relying on his alchemical, theological, and ancient-wisdom knowledge.

Newton's methodology is more complex than the generally accepted view of an axiomatic deductive or positivistic interpretation. Within Newton's method of resolution (analysis) and composition (synthesis), induction plays a crucial role. The first four "rules for the study of natural philosophy" in *Principia* have become emblematic to understand Newton's "experimental philosophy." In particular, the controversial Rule 4, which was added for *Principia*'s third edition (1726), states that "in experimental philosophy, propositions gathered from phenomena by induction should be considered either exactly or very nearly true notwithstanding any contrary hypothesis, until yet other phenomena make such propositions either more exact or liable to exceptions." Besides the method of analysis and synthesis, Newton developed his conception of a potentially open-ended process of successive approximation to reality similar to that of critical rationalism's verisimilitude.

The Scottish Enlightenment and Adam Smith

The Scottish Enlightenment rapidly assimilated Newton's discoveries and his methodology, spreading Newtonianism first in Britain and then in the Continent, where it was received more reluctantly. The Cartesian tradition could not accept the notion of a void, insisting on mechanisms and contact between bodies. Newton's notion of universal gravitation as a force operating universally and independently of any direct mechanical contact was difficult to accept in France. René Descartes had defined matter as an infinitely extended *plenum,* but Newton formulated his concept of universal gravitation operating in bodies *in vacuo*. G. W. Leibniz, the most capable representative of mechanical philosophy, who discovered calculus at the same time as Newton, argued that a contact mechanism was needed to explain gravity. And if Newton had to appeal to God in the *General Scholium* to explain gravity, his realism is explicit when he declares that gravity "simply exists."

According to Dugald Stewart, Adam Smith's lectures in moral philosophy at Glasgow University (1752–1764) were divided into four parts: Natural Theology, Ethics, Jurisprudence, and Political Economy (*expediency*). Based on it, Smith developed a system of social science that comprised ethics (*The Theory of Moral Sentiments,* 1759), economics (*An Inquiry Into the Nature and Causes of the Wealth of Nations,* 1776), and his unaccomplished project of jurisprudence.

Naturally, Adam Smith reflects the 18th-century admiration for Newton's discoveries. In his essay *The Principles Which Lead and Direct Philosophical Enquiries; Illustrated by the History of Astronomy,* the father of economics analyzes "the superior genius and sagacity of Sir Isaac Newton," which "made the most happy, and, we may now say, the greatest and most admirable improvement that was ever made in philosophy." In addition, paraphrasing Query 31 of *Opticks* for extending Newton's methodology to the realm of moral philosophy, Smith is reported to have lectured that "the Newtonian method is undoubtedly the most Philosophical, and in every science whether of Moralls or Naturall philosophy." This is textual evidence that Smith relied on Newton, but the crucial question is about the real nature of Smith's Newtonianism.

Based on some references to Newton and his system, and a narrow understanding of Smith's notion of human beings as a simple homo economicus, it has been generally accepted that Smith attempted to build his system of social science on a Newtonian basis. If the French early reception of Newton's legacy finally led to neoclassical and modern mainstream economics, Smith had a different interpretation of Newtonianism. The Scottish Enlightenment not only promoted Newtonianism but also understood its methodology. Therefore, Adam Smith was well aware that we approximate to the real nature of social science using a progressive

process of approximation. In this sense, the real legacy of Newton's methodology was developed by the Scottish Enlightenment. And Adam Smith, the father of economics, was a key inheritor of this tradition.

Leonidas Montes

See also Critical Rationalism; Individualism, Methodological; Mill and the Moral Sciences; Popper's Philosophy of Science; Scottish Enlightenment: Influence on the Social Sciences

Further Readings

Cremaschi, S. (1989). Adam Smith: Skeptical Newtonianism, disenchanted republicanism, and the birth of social science. In M. Dascal & O. Gruengard (Eds.), *Knowledge and politics: Case studies in the relationship between epistemology and political philosophy* (pp. 83–110). Boulder, CO: Westview Press.

Fleischacker, S. (2004). *On Adam Smith's Wealth of Nations: A philosophical companion*. Princeton NJ: Princeton University Press.

Hetherington, N. S. (1983). Isaac Newton's influence on Adam Smith's natural laws in economics. *Journal of the History of Ideas, 44*(3), 497–505.

Montes, L. (2003). Smith and Newton: Some methodological issues concerning general economic equilibrium theory. *Cambridge Journal of Economics, 27*(5), 723–747.

Montes, L. (2008). Newton's real influence on Adam Smith and its context. *Cambridge Journal of Economics, 32*(4), 555–576.

Montes, L. (2013). Newtonianism and Adam Smith. In C. Berry (Ed.), *The Oxford handbook of Adam Smith* (pp. 36–53). Oxford, England: Oxford University Press.

Newton, I. (1979). *Opticks: Or, a treatise of the reflections, refractions, inflections and colours of light*. London, England: William Innys. (Original work published 1704)

Newton, I. (1999). *Mathematical principles of natural philosophy* (I. B. Cohen & A. Whitman, Eds.). Berkeley: University of California Press. (Original work published 1687)

Smith, A. (1981). *An inquiry into the nature and causes of the wealth of nations* (R. H. Campbell & A. S. Skinner, Eds.). Indianapolis, IN: Liberty Fund. (Original work published 1776)

Smith, A. (1982). *Essays on philosophical subjects* (W. P. D. Wightman & J. C. Bryce, Eds.). Indianapolis, IN: Liberty Fund.

Smith, A. (1984). *The theory of moral sentiments* (D. D. Raphael & A. A. Macfie, Eds.). Indianapolis, IN: Liberty Fund. (Original work published 1759)

Smith, A. (1985). *Lectures on rhetoric and belles lettres* (J. C. Bryce, Ed.). Indianapolis, IN: Liberty Fund.

Nihilism

Nihilism is derived from the Latin term *nihil,* meaning nothing. At its core, nihilism is the claim that existence has no meaning. This entry charts the history of the notion of nihilism and shows its importance both for philosophy and for social inquiry.

The most influential version of nihilism is *existential nihilism,* the claim that there are no ultimate values. Ultimate values are those values that give life, and existence itself, meaning. The claim that existence is meaningless is sometimes taken to be the central premise of existentialism of this sort. The term *nihilism* can also be used in an *epistemological* and a *moral* sense. Epistemological nihilism is a form of radical scepticism that denies that anything can be known. Moral nihilism is the metaphysical claim that there are no objective moral values. The nihilistic claim that life has no meaning is often twinned with one of two normative responses: either that one should withdraw from life and active engagement in it or that one should defy life by adopting a form of violent protestation against its absurdity, as suggested in the French existentialist writer Albert Camus's *The Myth of Sisyphus*.

At the end of the entry, we shall mention two forms of nihilism in recent analytic philosophy, *mereological nihilism* and *metaphysical nihilism*.

The concept of nihilism plays a prominent role in many philosophical and literary works from the middle of the 19th century onward and may be regarded as one of the distinguishing traits of modernity. In particular, nihilism is often regarded as one of the perceived effects of secularization on European culture. Nihilism's prominence in social thought owes as much to literary and cultural criticism as it does to systematic theorizing. This in part explains the concept's heterogeneity.

History

The German philosopher F. H. Jacobi, in an open letter to J. G. Fichte in 1799, claims that Fichte's philosophy of Absolute Idealism is "nihilistic" because it denies the existence of anything outside the transcendental Ego. As God (according to Jacobi) cannot be posited by Fichte's "I," Fichte's philosophy denies God and is therefore a kind of nihilism. Jacobi concedes that Fichte's system is intellectually rigorous; his concern is with the potentially deleterious

effects that belief in such a philosophical system may have on its proponents. His critique of nihilism as reason pushed to destructive limits was a recurring theme in its subsequent development. However, the term *nihilism* really gained widespread recognition through its use in the Russian author Ivan Turgenev's *Fathers and Sons*. Bazarov, the main character of the novel, exemplifies an extreme form of critical positivism (scientism), subjecting all beliefs and received opinion to critical scrutiny and doubt. Turgenev's ambivalence toward Bazarov is partially determined by his reaction to the Russian anarchist and Young Hegelian thinker Mikhail Bakunin, whom Turgenev met in Berlin in 1840. Bakunin had attained great notoriety for extolling creativity through destruction. The perceived link between nihilism and the social and political movement of anarchism (and to some extent socialism) was strengthened by another famous Russian, Fyodor Dostoevsky, whose work *The Devils* (1872) drew inspiration from the activities of the revolutionary nihilist Sergei Nechaev. Nechaev's notorious *Catechism of a Revolutionary,* published in 1869, called for the complete destruction of both the state and society. Peter Verkhovensky, the main character of *The Devils,* participates in various assassinations and other acts of political violence. Dostoevsky's critical appraisal of the nihilistic spirit as bound up with violent and destructive acts of rebellion against society had a crucial impact on the development of Friedrich Nietzsche's understanding of nihilism.

Nietzsche

Friedrich Nietzsche (1844–1900) is the most influential figure in the subsequent understanding and elaboration of nihilism at the end of the 19th and for much of the 20th century. Despite the fact that Nietzsche published very little on nihilism, many of his numerous notebook entries dealing with nihilism appeared in a posthumously published work, *The Will to Power,* put together from his notes by his sister and others. This volume proved extremely influential in the subsequent understanding of both nihilism and Nietzsche's own philosophy.

For Nietzsche, modern nihilism is the absence of all overarching (ultimate) values and is the inevitable outcome of a thorough recognition of the death of (the belief in) God. For Nietzsche, modern moral systems such as Kantian deontology, utilitarianism, socialism, and Marxism merely carried on what he called the Judaeo-Christian "morality of compassion" in a secular form. They, like Christianity, are moralities of compassion in that they are essentially concerned with alleviating the suffering of humanity. Nietzsche argued that without the authority of God's command these moralities of compassion will eventually lose their grip on us and we, modern Europeans, will be left without any overarching values. This may aptly be called the "nihilism of disorientation." It is the kind of disorientation that has often been noted (e.g., by Sigmund Freud) to accompany the great conceptual revolutions instituted through the discoveries of Copernicus (the earth is not the center of the universe), Charles Darwin (humankind is not the product of, and hence the center of, divine creation), and Freud (man's conscious ego is not his ruler and the center of his being). It is this disorientation, the lack of ultimate overarching, centring values, that Nietzsche presciently pronounces as "the history of next two centuries." The emphasis on overarching values is crucial, for, as Nietzsche himself claims, life inevitably involves values. For instance, in reading this essay you are expressing a value; you are, at the moment, valuing reading it above, say, watching television. In his *Thus Spoke Zarathustra,* Nietzsche paints a vivid picture of the "last men," who value their families, their careers, and their small happiness but ask, "What is a star?" and blink. If nihilism really involved the absence of all values, as opposed to only the absence of overarching values, it would be incompatible with life (but see Nietzsche on affective nihilism, below). The basic Nietzschean argument that life inevitably involves values is that life itself is nothing but a collection of drives, and each drive, through its aims, assigns instrumental values to things, according to the thing's ability to satisfy the drive's aim. Thus, our hunger drive places a high value on food, say the croissant before us, and little value on, for instance, the pretty curtains in our study. What the nihilist lacks are overarching values that give meaning and provide a narrative structure to all existence. This Nietzschean idea was influential in works such as Oswald Spengler's *The Decline of the West.* The post-Christian West, "disenchanted" in Max Weber's words, is arguably a culture in decline as its members have lost belief in the central organizing values that fortified and gave vitality to their culture.

However, beyond this modern nihilism of disorientation, Nietzsche recognizes a deeper form of nihilism. This is evidenced by his repeated claim that Christianity itself is nihilistic. The Christian after all does not suffer from disorientation. Existence for him has great value as it is God's creation and, for the righteous, a gateway to the infinitely valuable external bliss of heavenly existence. Nietzsche argues, however, that this very proclamation of the value of the world to come serves to disvalue this, our worldly, existence. Christianity tells us that our natural worldly inclinations, our sexual drives, and our drives for power and dominance are affronts to God and need to be repressed if not extirpated if we are to gain entry to heaven. Now if we, like Nietzsche, equate life with our drives, we will see Christianity's rejection of the drives as a turn against life itself, what Nietzsche calls "the will turned against life." For Nietzsche, the desire to quieten the drives is in fact the desire to be nothing, to not exist, akin to what Freud would later call the *death drive*.

This phenomenon of the will turned against life we might call *affective nihilism*. Unlike the nihilism of disorientation, it does not involve an endorsement of some proposition, for instance, the proposition that there are no ultimate values. It involves a psychological component: wholesale repression of the drives. This captures the *nihil* (nothing) of nihilism, since the drive to suppress or extirpate the drives is a kind of drive to nothingness; for, as we have observed, for Nietzsche, life is itself a collection of drives; thus, to eliminate the drives is to eliminate life itself. When Nietzsche heralds nihilism as the history of Europe for the next 200 years, what he is prophesying is that the affective nihilism of Christianity will come to be consciously expressed in the embracing of claims such as that of the nihilism of disorientation—that there are no ultimate values.

Other Senses of Nihilism

As well as its existential sense, nihilism can also be used in an *epistemological* and a *moral* sense.

Epistemological nihilism is a form of radical skepticism that denies that anything can be known. Taken literally, this appears to be paradoxical; if literally *nothing* can be known, then presumably this includes the claim that nothing can be known. So if epistemological nihilism is true, we cannot know that it is true. In the social sciences, epistemological nihilism is often closely associated with extreme forms of relativism, for instance, the claim that truth is always relative to a theory. Such claims face a similar philosophical problem, namely, that they are self-undermining. As the old joke has it, "Relativism is not true, relative to my theory." A common, though disputed, charge (as leveled, e.g., by Jürgen Habermas) against philosophical postmodernism (as in Jacques Derrida), with its celebrated "incredulity toward meta-narratives" (as Jean-François Lyotard urged), is that it leads to relativism and nihilism.

Moral nihilism is the claim that objective moral values do not exist. One version of this view was given by the 20th-century Australian philosopher John Mackie, who argued that moral properties do not exist. The world as revealed to us by science does not give any grounds for attributing such properties to things, despite what our moral language would suggest. Ethical language is thus in systematic error, according to Mackie, and hence this variety of moral nihilism is called an *error theory*.

Finally, in contemporary discussions in metaphysics, there have been two areas in which the notion of nihilism has been employed but in a different way from the ones we have presented so far. Whereas the first of these metaphysical theses is explicitly non-intuitive, the second is not. First, there is the thesis of *mereological nihilism,* advanced by philosophers who reject the existence of any proper parts; that is, it amounts to the thesis that there are no composite objects or, in other words, objects with proper parts do not exist. Advocates of mereological nihilism assert that there exist only the smallest building blocks of reality, *simples,* which, however are, and remain, independent, forming no composite wholes. So strictly speaking, there are no composite things, such as, for instance, houses, but only material simples (subatomic particles) "arranged housewise." Some nihilists of this sort, like Peter van Inwagen, admit living organisms as the only exception, allowing them to be composite wholes of parts that contribute to the organism's life. Second, there is the (quite intuitive) thesis of *metaphysical nihilism,* according to which it is possible that there could be no concrete object or that it is a possibility that nothing concrete existed. The thesis asserts a metaphysical possibility; it does not claim that there are no concrete things. Metaphysical nihilists support the thesis that an empty world is a possibility

by an argument, the so-called subtraction argument (originally put forward by Thomas Baldwin).

Ken Gemes and Chris Sykes

See also Normativity; Postmodernism; Relativisms and Their Ontologies

Further Readings

Baldwin, T. (1996). There might be nothing. *Analysis, 56,* 231–238.

Camus, A. (1975). *The myth of Sisyphus* (J. O'Brien, Trans.). London, England: Penguin.

Gillespie, M. A. (1996). *Nihilism before Nietzsche.* Chicago, IL: University of Chicago Press.

Lowe, E. J. (2002). Metaphysical nihilism and the subtraction argument. *Analysis, 62,* 62–73.

Mackie, J. L. (1977). *Ethics: Inventing right and wrong.* London, England: Penguin.

Nagel, T. (1979). The absurd. In *Mortal questions* (pp. 11–23). Cambridge, England: Cambridge University Press.

Nietzsche, F. (1967). *The will to power* (W. Kaufmann, Trans.). New York, NY: Random House.

Reginster, B. (2006). *The affirmation of life: Nietzsche on overcoming nihilism.* Boston, MA: Harvard University Press.

Rosen, S. (1969). *Nihilism: A philosophical essay.* New Haven, CT: Yale University Press.

van Inwagen, P. (1990). *Material beings.* Ithaca, NY: Cornell University Press.

Weller, S. (2001). *Modernism and nihilism.* London, England: Palgrave Macmillan.

Nonconceptual Content

The issue of nonconceptual content revolves around two related questions. First, are there mental or cognitive states that represent some entity, without the subject of those states having to possess and/or employ any concepts of that entity? Second, what is the nature of the contents that such cognitive states possess? Answering either question is extremely difficult. For one thing, *nonconceptual content* designates something in terms of what it is not, and there may be many distinct kinds of content that conceptual content is not. Just as "nonhoofed animal" does not designate a theoretically interesting biological kind, there is no guarantee that "nonconceptual content" designates a theoretically interesting psychological kind. For another, any argument for the existence of states with nonconceptual content, and any account of their nature, depends on a conception of conceptual content. Unlike the property of being a hoofed animal, there is a wide range of opinion regarding what conceptual content is and under what conditions a subject possesses it; and the importance and the plausibility of the claim that nonconceptual content exists depends on just how robust or deflationary one's notion of conceptual content is.

Content

As the term is used today, the claim that a mental state "has content" often just amounts to the claim that it represents something as being a certain way. Historically, this was not so. Bertrand Russell, criticizing the views of Alexius Meinong—and by extension Kasimir Twardowski—explicitly denied the existence of mental content. In contemporary discussions, a distinction between content and object is often not explicitly drawn, so whether the "content" that a mental state "has" is the object represented or something that itself represents that object—such as a Fregean thought—is an issue on which there is some disagreement. Some authors appear to use the term *content* for both sorts of things. For instance, Christopher Peacocke uses the phrase "the content of experience" to refer not only to the objects, properties, and relations perceived but also to the ways they are perceived. Some philosophers use the term *content* to designate what a mental state represents. Other philosophers expressly use the term *content* for the way or manner in which an object is represented. As David Woodruff Smith expresses it, the content is "the 'mode' of presentation"—that is, the conceptual or presentational structure of the experience.

The ambiguity of the term *content* has a significant impact on the shape of the debate. For those who identify a mental state's content with its object—what it represents—the claim that certain mental states have nonconceptual content can be taken to mean that such mental states represent different sorts of objects than do states with conceptual content, such as beliefs. This position is the "content view," or *absolute nonconceptualism.* It can also be taken to mean that nonconceptual and conceptual states differ not in their objects but qua states. This is

the "state view," or *relative nonconceptualism,* and is compatible with the claim that conceptual and nonconceptual states can represent the same kinds of objects. For those who distinguish a mental state's content from its object, the claim that some mental states have nonconceptual content is also compatible with the claim that those states represent the same objects as do states with conceptual content. All that follows is that they represent those objects in a different manner. So, for instance, one might explain the difference between thinking about a tree and seeing a tree as a difference not in the tree but in the contents of the respective states. On any version of nonconceptualism, there exist mental states such that the subject of such a state need not possess or exercise concepts of the objects (or "contents") the state represents in order to be in that state.

What sorts of states might have nonconceptual content? The most commonly discussed candidates are (a) perceptual states, (b) the states of animals and prelinguistic humans, and (c) subpersonal cognitive states.

Perceptual States

Nonconceptualists have specified a number of features that, allegedly, belong to perceptual states and that, allegedly, do not belong to conceptual states. The most widely discussed feature in the literature is the fineness of grain of perceptual states. For example, a table may exhibit a variety of shades of brown that are perceived, while the perceiver has no names for them. The argument, then, is that if we can be conscious of certain objects or features but lack concepts of those objects or features, then the states in which we are conscious of them must have nonconceptual content. Since perceptual states sometimes make us aware of certain objects or features, such as determinate colors, for which we lack concepts, those states have nonconceptual content.

The standard conceptualist response is to claim that we do have concepts fine-grained enough to characterize such properties, namely, demonstrative concepts, and that such demonstrative concepts are exercised in perception. As John McDowell (1994) put it,

> In the throes of an experience of the kind that putatively transcends one's conceptual powers . . . one can give linguistic expression to a concept that is exactly as fine-grained as the experience, by uttering a phrase like "that shade." (pp. 56–57)

One nonconceptualist response to this position is that having the concept *C* entails having the ability to reidentify the object or property falling under *C,* but perception discrimination of a property does not require having the ability to reidentify it. Another response is that the perceptual experience of an object is what causally explains one's ability to refer to it demonstratively, in which case exercising the demonstrative concept cannot partially constitute one's perceptual experience of the object.

Many advocates of nonconceptualism, including some who reject the demonstrative strategy, express reservations about the fineness-of-grain objection, at least if that objection amounts to the claim that perceptual experiences represent more determinate properties than can be represented conceptually. There is another way of understanding the fineness of grain of experiences, however, according to which it is not that the objects of experience are more fine-grained or determinate than those of conceptual states. Rather, it is that the states represent objects and properties in a more fine-grained *way.* According to Sean Kelly, the nature of one's experience of color is partially determined by the lighting conditions in which it is perceived—the same color can appear differently in different contexts without appearing to be different. Since a demonstrative such as "that color" does not distinguish between the same color as experienced in different conditions, the experience has a sort of content that cannot be captured conceptually. Chistopher Peacocke makes the related point that the visual perception of a square differs from that of a regular diamond, despite their having identical objects and despite the fact that the demonstrative "that shape" would, if employed in both experiences, refer to the same shape. (The French phenomenologist Maurice Merleau-Ponty spoke, in the same vein, of the "paradox of symmetrical objects.") As we saw above regarding how the "content" of experience should be construed, Peacock insists that when we describe cases of fine-grained phenomenology, we employ the "notion of the *way*" a property or a relation is given in experience.

Another feature of perceptual states cited by nonconceptualists is that their contents are context and object dependent. The phenomenon of perceptual constancy shows that a given property

(shape, color, size, etc.) can appear to be the same in experiences that are phenomenologically different. What accounts for this is that perceived properties are always presented in a wider context. Kelly also argues that perception does not reveal abstract properties but "dependent aspect[s]" of the perceived objects. The blue of a woolly blue carpet and the blue of a shiny steel ball might be the same, but they show up differently in perceptual experience.

A related argument that perceptual states have nonconceptual content is that conceptual contents can be entertained when one is not confronted with the objects or states of affairs those contents represent, whereas the contents of perceptual states cannot. Conceptual content, as Susan Hurley said, is not "context-bound." Edmund Husserl had also argued that perception is not a "carrier of meaning," since on the basis of exactly the same perceptual experience one can think, "There flies a bird!" "That is black," and countless other propositions. And one can think, "There flies a bird!" while one's perceptual experience of the bird varies, and even when one is not perceiving any birds at all. More generally, for any perceptual experience *E* and any conceptual content *p,* undergoing *E* does not entail that one is entertaining *p,* and entertaining *p* does not entail that one is undergoing *E*.

Another argument that perceptual experiences have nonconceptual content is that their contents are not structured in the way those of conceptual states are. As Jerry Fodor argues, the concepts "John," "Mary," and "loves Mary" are *constituents* of the thought that John loves Mary. "John loves," however, is a *part* rather than a constituent, since it does not figure in the "canonical decomposition" of the thought that "John loves Mary." Perceptual contents, however, are iconic representations, which do not have any canonical decomposition. Such representations have parts but not constituents. Furthermore, according to Gareth Evans's Generality Constraint, a thinker capable of thinking both *Fa* and *Gb* is also capable of thinking *Fb* and *Ga,* provided that doing so does not involve any category mistakes. Perceptual contents, however, cannot be combined in this way. One might perceive that a cat is running and that a building is tall, but one cannot thereby perceive that the building is running.

Other advocates of nonconceptual content, finally, have emphasized the intimate connection between perception and action. Conceptually, one might present one's speed on a motorcycle as "55 miles per hour." That, however, is neither necessary nor sufficient for the sort of "embodied and environmental" knowledge one possesses while actually going that fast. Furthermore, nonconceptualists have argued that there is a constitutive connection between perceptual states and actions, while connections between conceptual states and actions are contingent.

All of these arguments depend on rather substantive claims about the nature of conceptual contents and their possession conditions, and it is open to the conceptualist to dispute them. Alva Noë (2004), for instance, writes,

> Perceptual experience presents the world as being this way or that; to have experience, therefore, one must be able to appreciate how the experience presents things as being. But this is just to say that one must have concepts of the presented features and states of affairs. (p. 183)

If employing concepts is just a matter of being able to represent something as being a certain way, then plainly none of the arguments for nonconceptual content above are sound.

Conceptualists, for their part, have defended their position in a variety of ways. One of the most notable arguments, by McDowell or Bill Brewer, for instance, is that since only states with conceptual content can justify beliefs and since perceptual states plainly do justify beliefs, experiences must have conceptual content. Against this, Tyler Burge has criticized such views for overintellectualizing perception and for treating perceptual belief as a form of reasoning, and Walter Hopp has argued that the special epistemic role of experience cannot be accounted for in any conceptualist view.

Animals and Children

Another argument for nonconceptual content is that animals and children are aware, and consciously aware, of objects in the world but lack concepts of many, most, or even all of the objects of which they are conscious. Sandfish, to use Andy Clark's example, detects vibrations caused by its insect prey. But it does not represent the features and occupants of its environment by employing the concepts we use to describe them, such as "vibration" or "insect." Furthermore, animals need not satisfy the Generality Constraint. A frog might represent that there is a

fly over there, without being able to be in any other sorts of states directed at flies. Similar remarks go for prelinguistic humans.

Conceptualists counter such arguments either by denying the theory of concept possession underlying them or by disputing the claims made about animals and children. Alva Noë argues that while animals surely do perceive and act in their environments, they also seem to demonstrate a number of conceptual and inferential capacities differing from our own only in degree. A monkey can recognize another "as of high status," for instance. That we treat our conceptual abilities as different in kind from those of animals is likely due to the fact that we have a much more exalted conception of our own conceptual abilities. When we examine how we recognize something as red, or an argument as invalid, we find that we can apply the relevant concepts without any ability to articulate why or how we can and that our abilities, too, are highly context-bound.

McDowell (1994a), by contrast, does not lower the bar on concept possession but insists that our conceptual and perceptual abilities differ from those of animals in kind. We do share something in common with animals. However, this does not mean that we can "isolate what we have in common with them by stripping off what is special about us" (p. 64), leaving perceptual states as a common factor in the lives of both persons and animals. Drawing on the hermeneutics philosopher Hans-Georg Gadamer, McDowell claims that animals inhabit an environment (Umwelt), while we inhabit a world (Welt). Animals are merely sentient; solely biological imperatives structure their lives rather than deciding what to do and think. Because of this, we should not ascribe full-fledged subjectivity to animals.

Subpersonal States

Finally, some philosophers have argued that subpersonal representational states have nonconceptual content. The argument is that such states have content but since the subject of those states need not possess the concepts required to specify that content, the content is nonconceptual. To borrow an example from Jose Bermúdez, David Marr's theory of vision involves a subpersonal, unconscious 2½-D sketch that represents depth and orientation. But the subject of such a state need not possess the concept "2½-D sketch," or indeed any of the concepts that figure in Marr's complicated theory. Bermúdez argues that such subpersonal states have *content* insofar as they serve a role in explaining behavior, can be integrated with other cognitive processes, are compositionally structured, and can possibly misrepresent their objects.

A conceptualist might argue that such states do not bear intentional content at all but simply manipulate syntactic items. McDowell, for instance, argues that subpersonal systems merely have "as-if" content and are "syntactic" rather than "semantic" engines. Another possible response is that the argument relies upon a flawed conception of conceptual content. An alternative view is that a content is conceptual just in case it is composed of concepts, and Bermúdez's arguments do not establish that the contents of subpersonal states are not composed of concepts.

Conclusion

Clearly, the debate over nonconceptual content is far from settled. There is no consensus over just what contents are, what concepts are, and what their possession conditions are. There are, in addition, substantive and controversial philosophical, phenomenological, and psychological premises in almost every argument for or against nonconceptual content. What should be clear, however, is that the current debate has generated a great deal of important insights into the nature of a wide variety of mental states.

Walter Hopp

See also Cognitive Phenomenology; Concepts; Consciousness; Embodied Cognition; Social Perception

Further Readings

Bermúdez, J. L. (2002). Nonconceptual content: From perceptual experience to subpersonal computational states. In Y. H. Gunther (Ed.), *Essays on nonconceptual content* (pp. 183–216). Cambridge: MIT Press.

Bermúdez, J. L. (2007). What is at stake in the debate on nonconceptual content? *Philosophical Perspectives, 21,* 55–72.

Brewer, B. (1999). *Perception and reason.* Oxford, England: Clarendon Press.

Brewer, B. (2005). Does perceptual experience have conceptual content? In M. Steup & E. Sosa (Eds.), *Contemporary debates in epistemology* (pp. 217–250). Malden, MA: Blackwell.

Burge, T. (2003). Perceptual entitlement. *Philosophy and Phenomenological Research, 67,* 503–548.
Clark, A. (2002). Connectionism and cognitive flexibility. In Y. H. Gunther (Ed.), *Essays on nonconceptual content* (pp. 165–181). Cambridge: MIT Press.
Crane, T. (1992). The nonconceptual content of experience. In T. Crane (Ed.), *The contents of experience* (pp. 136–157). Cambridge, England: Cambridge University Press.
Cussins, A. (2003). Content, conceptual content, and nonconceptual content. In Y. H. Gunther (Ed.), *Essays on nonconceptual content* (pp. 133–163). Cambridge: MIT Press.
Fodor, J. (2007). The revenge of the given. In B. P. McLaughlin and J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (pp. 117–138). Malden, MA: Blackwell.
Hopp, W. (2009). Conceptualism and the myth of the given. *European Journal of Philosophy, 17,* 363–385.
Hopp, W. (2011). *Perception and knowledge: A phenomenological account.* Cambridge, England: Cambridge University Press.
Hurley, S. L. (1998). *Consciousness in action.* Cambridge, MA: Harvard University Press.
Husserl, E. (1970). *Logical investigations* (J. N. Findlay, Trans.). London, England: Routledge & Kegan Paul.
Kelly, S. D. (2001a). Demonstrative concepts and experience. *Philosophical Review, 110,* 397–420.
Kelly, S. D. (2001b). The non-conceptual content of perceptual experience: Situation dependence and fineness of grain. *Philosophy and Phenomenological Research, 62,* 601–608.
McDowell, J. (1994a). *Mind and world.* Cambridge, MA: Harvard University Press.
McDowell, J. (1994b). The content of perceptual experience. *Philosophical Quarterly, 44,* 190–205.
Merleau-Ponty, M. (1962). *Phenomenology of perception* (C. Smith, Trans.). London, England: Routledge & Kegan Paul.
Noë, A. (1999). Thought and experience. *American Philosophical Quarterly, 36,* 257–265.
Noë, A. (2004). *Action in perception.* Cambridge: MIT Press.
Peacocke, C. (2001a). Does perception have a nonconceptual content? *Journal of Philosophy, 98,* 239–264.
Peacocke, C. (2001b). Phenomenology and nonconceptual content. *Philosophy and Phenomenological Research, 62,* 609–615.
Prinz, J. (2002). *Furnishing the mind: Concepts and their perceptual basis.* Cambridge: MIT Press.
Russell, B. (1984). *Theory of knowledge: The 1913 manuscript* (E. R. Eames in collaboration with K. Blackwell, Eds.). London, England: George Allen & Unwin.
Smith, A. D. (2002). *The problem of perception.* Cambridge, MA: Harvard University Press.
Smith, D. W. (1989). *The circle of acquaintance.* Dordrecht, Netherlands: Kluwer.
Speaks, J. (2005). Is there a problem about nonconceptual content? *Philosophical Review, 114,* 359–398.
Twardowski, K. (1977). *On the content and object of presentations* (R. Grossmann, Trans.). The Hague, Netherlands: Martinus Nijhoff.

Norbert Elias: Process of Civilization and Theory of Sciences

On the Process of Civilization (alternatively and mostly known by the misleading title *The Civilizing Process*) is the English title of the most famous book, *Über den Prozess der Zivilisation,* written by the German-born sociologist Norbert Elias (1897–1990). In this book, Elias investigates in a non-normative way the interconnected development of manners and social structures in Europe from the end of the Middle Ages. The "process sociology" (alternatively, "figurational sociology") that his work inspired is characterized by a long-term historical approach, with an interest in the multiple aspects of social life. Crossing disciplinary borders, the study of social processes is also based on cross-cultural comparisons. Process sociology consequently challenges all forms of reification, notably the fictive methodological opposition between society and the individual that is at the foundation of individualist and holist approaches. In addition, the theory of knowledge advocated by Elias and his followers poses questions about the development of the social sciences.

The Book

Über den Prozess der Zivilisation was first published in German in Basel in 1939, when its author was in exile in London. Based on European writings on manners, the first volume examines the changes in the behavior of the upper classes in the West through the centuries, for example at table. Elias demonstrates that disgust and reserve are not innate

or invariant. Behavior and sensibility have been continuously shaped according to what was socially expected, till a point where modeling became unconscious for individuals. The second volume links the long-term social process toward civilization with the dynamics of state formation in Europe.

The Western civilizing process thus involves at the same time "psychogenesis" and "sociogenesis." The influence of Sigmund Freud is evident, but Elias is more interested in showing the sociohistorical nature of any superego formation process, by inquiring into how social constraint progressively mutates toward self-constraint. For instance, the end of the age of feudal warriors and the progressive establishment of the state's monopoly of the legitimate use of violence, along with the greater development of economic interdependencies, consolidated the constraints imposed on the aggressive drive.

Unnoticed in 1939, Elias's book and the idea of a "blind" and unplanned but directional process were later subjected to many criticisms, aggravated by the context of World War II and the consequent contestation of the very possibility of talking about "civilization." Nevertheless, Elias's enquiry had little to do with the ethnocentric and unilinear progress theories that marked the emerging social sciences in the 19th century, even though he recognized August Comte and Karl Marx in particular as pioneers of process theories. But the Eliasian approach has even less in common with the radical right-wing, antidemocratic and pessimistic visions that flourished in the early 20th century, such as Oswald Spengler's thesis expounded in *The Decline of the West*.

Process Sociology

Not only did Elias never suggest that there was only one, European, process of civilization, but he invited challenges to his model by comparing the Western process with that of other societies, such as Asian court societies. Moreover, he pointed out that civilization has to be considered as a *process* that is neither good nor bad. For example, civilization does not mean the disappearance of violence from social life but rather a shift toward state military institutions and into an inner struggle that takes place *inside* highly self-controlled individuals. Elias also showed that state formation unleashed increased violence in the relations *between* states.

Civilizing processes are plural, reversible, and often go hand in hand with "de-civilizing" processes. This idea is central in the essays Elias devoted to Germany and to the Nazi period. In his *Studies on the Germans,* he tried to advance a valid explanation of what he called "the breakdown of civilization," without giving up the principles of process sociology.

Norbert Elias's sociology often depicts itself as "reality congruent." It implies first of all an interest in how societies change to become what they actually are. It also forces the recognition that a society, or "figuration," is always a society "of individuals." It means that there is no society existing apart from networks of interconnected individuals. At the same time, there is no *homo clausus* or isolated individual preexisting social life, as people are irreducible to the "thinking statues" depicted by the classical philosophy from Descartes to Kant. Individuals "depend" on each other in the various aspects of their life and inherit their institutions, ideas, manners, behavior, and mode of thinking and feeling from precedent generations, although they continuously transform them. The different areas of social reality, from politics to sports or emotions, are indeed linked through individuals, instead of being separated, as they seem to be for the hyperspecialization that occurred in the social sciences and that Elias denounced, as well as "the retreat of sociologists into the present."

Reality congruence thus imposes a great interdisciplinary openness and a long-term historical approach, as some of the features of societies are transformed on the scale of centuries. Process sociology is far from according with a certain vision of figurations as stable or harmoniously evolving systems. While avoiding any kind of determinism, it rather seeks in the changing balances of power between groups the ways figurations are potentially changing in their various aspects.

Avoiding the structuralism and functionalism that dominated sociology in the second half of the 20th century, Elias did not really found a unified school of thought. Yet his work became particularly influential in the Netherlands and eventually in Britain, two countries where he taught and lived. In France, his sociology has renewed interest in a diachronic sociological approach, notably because of a certain proximity with the historians of the

Annales school and with Michel Foucault's themes. Pierre Bourdieu contributed too to popularize the concept of *habitus,* which Elias had used much earlier to describe embodied social knowledge.

Theory of Knowledge

Elias was quite suspicious of any form of idealism. He first gained a doctorate in philosophy, but he had a major dispute with his teacher, the Neo-Kantian Richard Hönigswald, about the Kantian idea that some a priori elements of thought exist that do not derive from experience but are permanent and universal. Process sociology is particularly concerned about balancing empirical observations against theory. In Elias's view, *what* sociology unveils (i.e., discovery) in a sense prevails over *how* it does it. Elias totally disagreed with Karl Popper. First, the sociologist contested the preeminence of methodology in itself. Second, his writings are very critical of the predominance of the methods imported from physics and the natural sciences. Given their historical character, processes that constitute social reality are in fact too complex to be comprehended by means of only quantitative and analytic tools, observation, and measurements. On that point, Eliasian process sociology inherits the questions raised by German epistemology, which, from Wilhelm Dilthey to Max Weber, feed a fundamental distinction between *Kulturwissenschaft* and *Naturwissenschaft,* as expressed by Heinrich Rickert in 1899.

Nevertheless, the process approach tries to avoid all reifying oppositions of that kind. Sociology is a particular standpoint focused on people living together, while psychology interests itself in people as individuals. And a human being also is, or has, a body, that is within the scope of biology. In other words, sciences are linked to each other through their common human subject. Not only do the humanities and social sciences depend on the natural sciences and physics, but the latter are also social products that historically vary—which made them subjects of investigation by the former.

This perspective of continua rather than dichotomies remains present in Elias's reflections on the dynamics of "involvement and detachment." The main issue here is that of the situation of the researcher, and it is irreducible to the problem of objectivity. The social scientist in particular is evidently part of the social life. Elias seeks to ask how this involvement varies and to what extent it is necessary to perform his role or, at the opposite, the extent to which it may blind him to certain observations. He thus enlarges Max Weber's question about the usefulness of a social scientist's commitment to certain social or political beliefs. Process sociology also proposes a reflection on how a certain amount of self-detachment can be reached. Finally, it reopens the debate on the conditions in which sociologists may, or should, face the problems of their times.

Florence Delmotte

See also Habitus; Holism, in the Social Sciences; Individualism, Methodological; *Naturwissenschaften* Versus *Geisteswissenschaften;* Philosophy of Sociology, History of; Popper's Philosophy of Science

Further Readings

Elias, N. (2007). *Involvement and detachment* (Collected Works, Vol. 8). Dublin, Ireland: University College Dublin Press.

Elias, N. (2009). *Essays I: On the sociology of knowledge and the sciences* (Collected Works, Vol. 14). Dublin, Ireland: University College Dublin Press.

Elias, N. (2010). *The society of individuals* (Collected Works, Vol. 10; E. Jephcott, Trans.). Dublin, Ireland: University College Dublin Press. (Original work published 1987 in German)

Elias, N. (2012a). *On the process of civilization* (Collected Works, Vol. 3; E. Jephcott, Trans.). Dublin, Ireland: University College Dublin Press. (Original work published 1939 in German; previous editions published under the title *The Civilizing Process*)

Elias, N. (2012b). *What is sociology?* (Collected Works, Vol. 5; G. Morissey, S. Mennell, & E. Jephcott, Trans.). Dublin, Ireland: University College Dublin Press. (Original work published 1970 in German)

Elias, N. (in press). *Studies on the Germans* (Collected Works, Vol. 11; E. Dunning & S. Mennell, Trans.). Dublin, Ireland: University College Dublin Press. (Original work published 1989 in German)

Kilminster, R. (2007). *Norbert Elias: Post-philosophical sociology.* London, England: Routledge.

Mennell, S. (1998). *Norbert Elias: An introduction.* Dublin, Ireland: University College Dublin Press.

Normativism Versus Realism

This entry introduces the positions of normativism and realism or descriptivism in explaining an agent's thoughts and actions and presents in detail the various versions of those positions as well as the arguments and counterarguments used by each. Normativism and realism are two very general, conflicting approaches to the explanation of what is thought and done.

Normativism seeks to explain by showing that the thought or action in question was normatively appropriate. It is thought that one explains better, more satisfactorily at least, if one exhibits the normative propriety of what is thought and chosen. Exhibiting rationality, for example, is commonly taken to be central to good explanation.

Realism/descriptivism seeks to explain by showing that the thought or action in question was the result of facts about the agents, societies, and environments, where such facts include regularities in the processes that are in play. Thus, at the level of individuals, one explains by understanding the cognitive processes in play—whether or not these are rational. Or one explains by understanding the conative commitments and preferences of the people involved. Of course, people are in various ways both rational and irrational—what is important in explanation is how they are, not how they should be.

Each approach has found adherents in the social sciences and among philosophers of the social sciences. Later in this entry, a related, but different, way of drawing the normativism/realism distinction will be presented.

Explanation

Influential thinkers such as the sociologist Max Weber have been interpreted as seeking explanations that work in both ways—finding rationality at the level of meanings yet uncovering causes that turn on mental processes, and allowing for some idealization on both counts. Normativism has constrained thinking in some paradigmatic social science. Thus, rational choice models have been influential in economics—commonly with little attention to the plausible limits of human rationality (or to the limited information commonly possessed by agents). On their face, such explanations apply normative models—decision-theoretic models of how folk ought to think and choose—to account for what is done and its aggregate consequences. Some political science and other related social sciences have followed suit. The influential philosopher Donald Davidson once insisted that rationalizing explanation is the only model of explanation with clear application to "the mental." Yet Davidson also insisted that explanations of why someone did or thought as they did could only be understood as a form of causal explanation, and causes are apparently real events in a real process. Covering-law models of causal explanation obviously incline one to realism—as reflected in Carl Hempel's classic discussion of covering laws in history. More recent general views on explanation within philosophy of science repudiate central elements of Hempel's deductive-nomological account, while yet generally supporting descriptivism.

A general idea shared by most approaches to explanation is the following: Explanations provide answers to why-questions—or (relatedly) to how-questions or what-if-things-had-been-different-questions. To explain why someone did something is to note what it is about them (or their conditions) that eventuated in their doing that rather than doing alternatives. Accordingly, to explain is to exhibit or allow one to appreciate a pattern of counterfactual dependencies. To understand the pattern of counterfactual dependencies is to understand what would have obtained had things been different with respect to the matters pointed to in one's explanation.

The conflict between realism and normativism then boils down to a question of whether, for actual thoughts and actions of actual humans, the pattern of counterfactual dependencies traces the nuances of rationality (or some similar normative matter) or, instead, follows the intricacies of actual cognitive processes and social processes, whether rational or not.

"Oughts"

It will be useful to draw a distinction between two conceptions of what is normatively appropriate—as those who suppose that one can explain by showing normative correctness may have either in mind. This allows for importantly different forms of normativism. On the one hand, there are *qualified oughts*, having to do with how people situated in a society

with certain norms ought to behave—among the *X*s (or as an *X*), one ought to *A*. On the other, there are *unqualified oughts,* thought to hold of any agent, *qua agent*—flatly: One (or all) ought to *A*. The qualified oughts would seem to be an interesting hybrid—as they would seem to be oughts that depend in part on descriptive matters holding within a group of people (their shared understandings, preferences, etc.) and in part on some unqualified oughts.

The distinction can be illustrated by thinking of how one thinks about correct moves in choice situations of the sort characterized by certain kinds of economic games. Consider the ultimatum game. In this two-person game, one player, the proposer, is conditionally provided a stake and must propose a division of that stake between the two players. The second player, the responder, can only accept or reject the proposal. If the proposal is accepted, the stake is divided accordingly. If it is rejected, neither player gets anything. Among folk in societies that have significant involvement with modern market economies, it is common to find proposals of 40% to 50% of the stake. Apparently, in those societies, there is a norm that prescribes roughly equal distributions. There is an idea in such societies that this is how one ought to play that game—perhaps provided that enough others also conform to the norm. There also seems to be a norm that the responder should punish proposers who offer 20% or less—they should decline the proposal at a cost to themselves as well as the proposer. This is how people in such societies tend to play such games—it was what, in those social contexts, is *counted as* how one ought to play. These patterns of play are pretty robust across the societies in which the ultimatum game was first studied (e.g., educated populations in the United States, Europe, Japan, and Indonesia).

However, when Joseph Henrich (and then others) ran ultimatum (and related economic) games in relatively isolated small-scale societies, they found much cross-cultural variation in play. Apparently, there are variations in the norms applied when encountering such games or situations. Somewhat provocatively, Henrich and associates distinguished Western, Educated, Industrialized, Rich, Democratic (WEIRD) societies, from others. It seems that what is normative for WEIRD people may not be normative for non-WEIRD people—as the norms for economic play may vary. Fitting to different settings and resources, different peoples have coordinated around different norms providing solutions to the choice situations they face. Arguably, we WEIRD folk ought to play as we commonly do, while the Machiguenga ought to play as they commonly do. Of course, in each social setting, there is a good deal of variation across members in the degree of norm conformity and norm violation. Normativists who envision explanation by exhibiting conformity with such qualified oughts might be termed *local normativists*.

On the other hand, there are those who imagine that there are objective oughts—holding for all people, no matter what their personal or social normative commitments. One might find such an unqualified normative principle lurking in or across the various social responses to the ultimatum game—as it might be argued that these different norms are satisfactory solutions to social problems faced in their respective settings. This is to envision a kind of universal game-theoretic normativity—and one might insist that this unqualified normativity is what makes for explanations. Indeed, some level of practical rationality and epistemic rationality commonly has been espoused by normativists as the unqualified normativity to be exhibited in explanations of thought and action. Normativists who envision explanations that work by exhibiting such unqualified oughts may be termed *imperial normativists*.

Imperial Normativism

Let us focus first on imperial normativism. On the epistemic side, it is said that all should accommodate base rates when making inferences about probabilities. But, famously, not everyone does, and perhaps no one does so consistently. Plausibly, perhaps, all should conform to socially salient norms that afford solutions to social choice problems, when one expects that others will conform. Again, not everyone does. Realists/descriptivists have a reasonably straightforward argument against imperial normativists that takes its departure from this mixed record of rationality and irrationality (or normative correctness and violations). Start with these uncontroversial points:

1. Sometimes folk do what they ought. Sometimes they do not. (Actually this is true, whether one uses qualified or unqualified oughts. So the argument rehearsed here will also have some force against local normativism.)

2. When folk do what they ought, there are facts about them—their actual training and resulting cognitive processes thus shaped—that are such that, in their environment or setting, tend to yield decisions/thoughts that (at least on the limited matter in question) conform to the relevant normative principles.
3. When they do not do what they ought, there are facts about them that are such that, in their environment or setting, tend to yield decisions/thoughts that (on the limited matter in question) fail to conform to the relevant normative principles.

Consider then the two classes of cases:

4. Obviously, when folk do not do what they ought, only the descriptive/realistic explanation has application. They do what they do, which is to fail to do what they ought, because of the facts about them that yield the normatively inappropriate result. The answers to why questions about their thought or action—and patterns of counterfactual dependency—all turn on the processes actually in play in them.
5. When folk do what they ought, it is no less true of them that the descriptive/realistic explanation of their action has application: In those cases, they do what they do because of the facts about them that yield at least these (limited range of) normatively appropriate results. The answers to why questions about their thought or action—and patterns of counterfactual dependency—all turn on the processes actually in play in them.

So, in either case, it is descriptive facts about the agents in their setting that explain what they do and think. These facts give rise to the patterns of counterfactual dependency the understanding of which makes for an explanatory answer to a why- (or a related) question.

6. The patterns of counterfactual dependency obtaining in any case are governed by the relevant descriptive facts—the actual processes actually in play—rather than by any general principle that the imperial normativist might adduce.

Understandings of cognitive/connotative processes, of background shaping of such cognitive processes, and of the resulting dispositions can enable one to answer why-, how-, and what-if-things-had-been-different questions. That some act or thought is flatly normatively appropriate cannot. Were one to attempt to explain what is done by noting that it is normatively appropriate in this unqualified sense—perhaps by saying flatly that it was rational—this would project a pattern of counterfactual dependencies that just does not obtain.

Here is a tractable toy illustration: Suppose that Agent A is given information about B. B was given a test for some condition, C; this test is 90% accurate, and B's test was positive for C. A then forms the belief that B has a 90% chance of having C. Suppose further that A's extant processes for working with probabilities are pretty minimal and that A would judge the same way were C a very rare condition or a very common condition. We may suppose that A was told (and believed) that there was no information to be had about how common Condition C was in the population. However, we have just stipulated that A is insensitive to base-rate information—such being A's extant processes. A would have neglected base-rate information had it been provided.

Of course Bayes's Theorem is a theorem of the probability calculus—the sort of principle that is generally loved by imperialist normativists. It says that such differences in base rates should be accommodated, at least when available. But, lacking such information, arguably, A's simple inference is rational in that it conforms to an application of Bayes's Theorem that is acceptable, given the ignorance of base rates stipulated.

To explain A's conclusion in terms of A's simple processes seems fitting—as the counterfactual dependencies obtaining here would be just those occasioned by the inferences that those processes would generate. Some of these inferences conform to Bayes's Theorem and would be normatively appropriate. Many do not. Understanding A's processes allows us to appreciate that, had A been told that C was a rare condition, A would still not have reached a different conclusion. Counterfactual dependencies follow actual processes, and what is explicable are the effects of actual processes.

In contrast, to explain A's simple inference as rational (and thus as fitting in light of Bayes's Theorem, taken as a putative explanatory principle of imperialist normativism) would entail that, had various information about base rates been provided

to A, A would have made correspondingly different inferences. But that is not so.

There is nothing special about this toy example. This problem for imperialist normativism will arise as long as (a) there are principles of practical or epistemic rationality expressing unqualified oughts and (b) people's dispositions deviate from what are taken to be such unqualified oughts. Whenever this obtains, and it seemingly is pervasive among actual cognitive systems, among actual creatures, the patterns of counterfactual dependency obtaining in any given case will follow the actual cognitive processes of flesh-and-blood agents, not the normative principles to which the imperialist normativist is committed. Bayes's Theorem simply served in the toy example as a possible illustration of one such principle of rationality. Provide a substitute if you prefer. The lesson will remain the same.

Local Normativism

Do social norms and local normativity help the normativist?

The kind of qualified normative principles called for by local normativism are social norms around which people have somewhat coordinated their actions or thinking. Thus, while members of diverse groups apparently respond to economic games differently, it seems that in each society their respective typical responses are informed by their respective social norms for the distribution of goods. WEIRD folk deploy a norm or set of norms that they have learned and adopted. Others deploy other norms. In each case, the character or content of a folk's norms may be largely implicit and difficult for folk to articulate. But among a folk, there is commonly some set of normative commitments that are learned and adopted. More or less coordinated around the application of these norms, people come to rely on each other, evaluate each other, and punish or reward each other in somewhat coordinated ways. Of course, the coordination, evaluation, conformity, and punishment are "uneven" or variable within the group.

Local normativism supposes that there is some sense in which people in such a setting *ought* to do as their norm requires. It is important that this is understood as a "real" normative ought and not some descriptive ersatz for oughts. Of course, such local oughts might be thought of as arising from the local application of some unqualified normative principles—such as principles of practical rationality. If so, then local normative principles amount to special-case applications of imperialist normativism. Local normativism would then face the objections to imperialist normativism mentioned above. If it is not understood in this fashion, it would seem to require that one be able to derive an "ought" from an "is": Local normativism seems committed to local oughts falling out of a descriptive state of affairs—the state of affairs that constitutes people's holding to some norm. In either case, local normativism itself faces daunting objections.

It is perhaps worth noting that normativists generally have resisted the idea that real oughts could arise merely out of some local descriptive facts-on-the-ground. The point is pressed by Richard Brandom, who insists that there must be a distinction between something being correct and its merely seeming correct. As Brandom argues, it does not help socialize the matter, as there is a distinction between something being right and people thinking, or acting as if, it is right. Without this difference, it is hard to see what normativism comes to or how it differs from a form of realism/descriptivism that emphasizes that one need attend to people's normative beliefs and attitudes. With this difference, it is hard to see what makes for the normative statuses that are envisioned—unless it is what the imperialist imagines. And even these can seem spooky. Such issues are pursued by Stephen Turner.

Social Norms

What it takes for there to be a social norm in some group can be understood in descriptive terms—for example, in terms of the preferences among some folk who have learned some action-directing content. For purposes of illustration, we can here draw on an account that seems as good as any presently on offer.

Cristina Bicchieri understands norms as rules that allow folk to coordinate behavior in the face of a class of choice situations (as solutions to some decision-theoretic game). On her usage, social norms are a special class of norms, having to do with situations in which individuals have some temptation to defect from the coordination (with "mix-motive games"). Descriptive norms are another class of norms, having to do with situations in which, once people have achieved a coordination by way

of the norm, there is no temptation to defect (with "coordination games"). In either case, the rules in question need not be articulable by agents, and how they are represented in individual agents is somewhat indeterminate. On Bicchieri's account of social norms, a rule is a social norm in a population facing a mixed-motive game just in case sufficiently many agents within that population (a) "know" the rule (to some approximation) and (b) have a conditional preference for conforming to it, provided (i) enough others conform to it and (ii) enough others have a tendency to evaluate performance according to the rules (and perhaps also to punish violations). A rule is a descriptive norm within a population just in case sufficiently many agents (a) know the rule (to some approximation) and (b) have a conditional preference for conforming to it provided enough others do. The extra demands on social norms are there to get over the mix of motives. In either case, provided one expects that enough others satisfy the conditions featured in one's conditional preferences, one likely conforms to the rule in question.

For our purposes, it is crucial to see that such characterizations of norms are wholly descriptive. It simply requires that a bunch of agents in the population (a) have some manner of mental representation of a behavioral rule and (b) have conditional preferences for conforming to that rule. Conformity then results when enough of those who hold the norm also have expectations for others conforming and evaluating according to that rule. One could be a skeptic about all normativity and yet have no difficulty applying this account of norms. Something along these descriptive lines is the "is" from which local normativism would need to derive its "oughts."

Realist/Descriptivist Responses and the Normativist Rejoinder

But the descriptivist/realist will certainly have reason to wonder what talk of normativity could add to the explanation afforded by a descriptive understanding offered by a descriptive account along the lines just sketched above. All patterns of counterfactual dependencies would seem to fall out of the cognitive representations and preferences just sketched—together with an understanding of the cognitive processes common with this set of agents. To the extent that one understands the content of the agents' representations of the norm, to the extent that one understands their conditional preferences, and their expectations for the satisfaction of those conditions (all matters to be described), and to the extent one understands the character or tendency of the agents' cognitive processes, one is wonderfully placed to understand why individuals do or think as they do and how they would act or think were they to be situated in different but related ways. Indeed, it is hard to see how one could be better placed to explain. Normativity, mysterious in its relation to these descriptive matters, seems to add nothing to one's ability to explain.

All the above discussion has presupposed one thing that the normativist will want to challenge: It has supposed that what is to be explained—the what is thought and done—is itself understood as descriptive matters of fact. This is at the heart of the disagreement between the normativist, exemplified by Brandom, and the descriptivist, exemplified by Turner. The normativist insists that some fitting topics of explanation have ineliminable normative elements. To appreciate what is at issue, at least to a first approximation, consider a case in which some agent, A, wants to go to the shop directly across the street from her, midblock. Suppose that A then walks down the street to the intersection and crosses there, before walking back up the other side of the street. In so doing, A conforms to the law prohibiting jaywalking in the area. And, indeed, we may suppose that A took the indicated route in part because she wanted to conform to the law and believed that the law prohibited jaywalking there. Of course, the descriptivist/realist can explain why A took the route described. Her desire to get to the shop, her understanding of its location relative to her and the block, her desire to comply with the law, and her belief about what the law prohibited, and an understanding of some rudimentary cognitive processes in A, should jointly afford the descriptivist an explanation. Indeed, the law in A's locality might not prohibit jaywalking, it would not matter—A would yet take the route indicated. The law may prohibit crossing at that intersection; again, it does not matter so long as A's beliefs are as indicated above. A may irrationally believe that she would be struck dead by the traffic gods, so that the explanation really involved irrationality, not rationality—again, this is fine with the descriptivist, who readily explains A's taking the route taken.

But the normativist will insist that there are things that the descriptivist cannot explain, and that one

wants explained—notably, why A conformed to the law. Of course, if "the law" is simply some descriptive matter, the descriptivist has no problems. Thus, if, crudely, the law in question is a matter of a social norm that a group of people have organized (with other social norms) to enforce, then the descriptivist can describe the relevant social norms and how A came to share in them, and the descriptivist can then explain A's conformity with the norm. However, some normativists insist that law—real law, and thus real conformity to law—requires a *legitimate* government and requires that a certain substantive principle be the result of actions by that government that are such as make a prescription or proscription into a law. They will insist that there is a difference between people—even all the people—thinking that such is a government and its being a government. They will insist that there is a difference between people—even all the people—thinking that something is the law of the land and its indeed being the law. This difference, they insist, makes a difference for the social sciences. To explain A's conforming to the law requires registering this difference, and it is a normative difference. The descriptivist/realist may counter that the normativists may not be up to their own challenge here.

David K. Henderson

See also Bayesianism, Recent Uses of; Cooperation, Cultural Evolution of; Cooperation/Coordination; Explanation, Theories of; Naturalism in Social Science; Normativity; Rationality and Social Explanation; Social Conventions; Social Norms

Further Readings

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms.* New York, NY: Cambridge University Press.

Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment.* Cambridge, MA: Harvard University Press.

Cummins, R. (1983). *The nature of psychological explanation.* Cambridge: MIT Press.

Davidson, D. (1980). *Essays on actions and events.* Oxford, England: Clarendon Press.

Davidson, D. (1984). *Inquiries into truth and interpretation.* Oxford, England: Clarendon Press.

Hempel, C. G. (1965). *The function of general laws in history. Aspects of scientific explanation and other essays in the philosophy of science.* New York, NY: Free Press.

Henderson, D. (2005). Norms, invariance, and explanatory relevance. *Philosophy of the Social Sciences, 35*(3), 324–338.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H. (Eds.). (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies.* Oxford, England: Oxford University Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). *The weirdest people in the world? Behavioral and Brain Sciences, 33*(2–3), 61–83.

Salmon, W. C. (1989). *Four decades of scientific explanation.* Minneapolis: University of Minnesota Press.

Turner, S. P. (2010). *Explaining the normative.* Oxford, England: Polity.

Woodward, J. (2000). Explanation and invariance in the special sciences. *British Journal for the Philosophy of Science, 51*(2), 197–254.

Normativity

Normativity is a property possessed by propositions. A proposition is normative just in case it is either an evaluative or a directive, or an appropriate function—such as negation, conjunction, or universal generalization—of evaluatives and directives. A proposition is an evaluative just in case a person who asserts it commends or "discommends" something; examples include the propositions that Alfred is a good person, that Barbara is a better parent than Carol, and that David is greedy. A proposition is a directive just in case a person who asserts it claims that someone or something is called on, or required, to behave in a certain way; examples include the propositions that Ellen should apologize to her brother, that Frank ought to help George, and that Harriet must stop hitting her sister.

1. It is intuitively very plausible that directives are in some way related to evaluatives. On many views, directives are analyzable into evaluatives. Objective consequentialists say that for it to be the case that Frank ought to help George is for it to be the case that the consequences of his doing so would be better than the consequences of his not doing so. Subjective consequentialists say that for it to be the case that Frank ought to help George is for it to be the case that Frank believes that the consequences

of his doing so would be better than the consequences of his not doing so. According to a more complex analysis: for it to be the case that Frank ought to help George is for it to be the case that if Frank knew what would happen if he did, and what would happen if he didn't, then he would be in a measure defective if he didn't.

2. The examples so far mentioned are all moral propositions, that is, propositions such that to assert them is to make a moral judgment. All moral propositions are normative. All aesthetic propositions are also normative, for they too are either evaluatives or directives, or appropriate functions of evaluatives and directives. (On some views, there are no true aesthetic directives, and thus all true aesthetic propositions are evaluatives or appropriate functions of evaluatives.)

But *are* there any such propositions? In virtue of finding it unclear what could make a moral proposition true, some moral philosophers conclude that there are no moral propositions. Thinking it equally unclear what could make an aesthetic proposition true, and taking it for granted that the moral and the aesthetic exhaust the normative, they conclude that there are no normative propositions at all. On their view—currently called *expressivism*—to say the words "Alfred is a good person" is merely to display a favorable attitude toward Alfred *qua* human being; to say the words "That statue is beautiful" is merely to display a favorable attitude toward the appearance of the statue referred to by *That statue;* to say the words "Harriet must stop hitting her sister" is merely to display an unfavorable attitude toward Harriet's hitting her sister. A substantial literature has been devoted to efforts to refute or refine that view.

3. Assuming that *expressivism* is false, and thus that there are such things as normative propositions, it is intuitively plausible that the true ones divide into (1) those that are made true by (1a) formal social arrangements or (1b) informal social conventions, and would not have been true if there had been no such arrangements or conventions, and (2) those that are true whether or not there are formal social arrangements or informal social conventions that also make them true.

Consider moral propositions. Suppose Alfred plans to kill Bert, though Bert has done no harm and doesn't plan to do harm. Suppose Carol therefore asserts that Alfred ought not kill Bert. Then she asserts a moral proposition of kind (2): it is true, and its truth is overdetermined since it is also made true by a law, thus by a formal social arrangement, but would have been true even if there had been no such law. (The question what makes moral propositions of this kind true lies at the heart of moral theory.)

Suppose, by contrast, that we tell people that when in America, they ought to drive on the right. Then we typically assert a moral proposition of kind (1a), which is made true only by a law—it wouldn't have been true if there had been no such law. On many views, the same holds of the directives that are made true by possession of property rights, since property rights are themselves generated only by law. (Legal and political theorists have focused on the question what makes moral propositions of this kind true.)

Two further points about moral propositions of kind (1a) should be mentioned. First, government is not the only source of formal social arrangements that make moral propositions true. For example, all private clubs have explicit rules governing their members, such as that members must pay their dues on time. Also, people very often explicitly contract or promise to do things, thereby making it the case that they ought to.

Second, while a formal social arrangement may clearly make some moral propositions true, it may need interpretation if it is to become clear whether it makes others true, and interpreting it may require knowledge of the informal social conventions in force in the relevant community. Suppose, for example, that Alfred contracted to deliver a cow to Alice. Then he clearly breaks their contract if he doesn't deliver a cow. The informal social conventions in force in their community may make it also true that he breaks their contract if he delivers a dead cow, even though he delivers a cow and the contract doesn't explicitly say "*live* cow."

Suppose, finally, that there was a line at the bus-stand, and that for no good reason, newcomer Alfred went to the head of the line. If Bert therefore asserts that Alfred ought not have done so, then we can suppose that he asserts a moral proposition of kind (1b), made true only by the existence of an informal social convention to the effect that newcomers go to the end of the line at bus-stands. (Social theorists have been particularly interested in the question what generates informal social

conventions, and why people act in accord with them when they do.)

4. The normative is not limited to the moral and the aesthetic, however. Suppose Alfred volunteered to set the table for dinner, and Alice therefore asserts a normative proposition to the effect that he ought to put the forks on the left of the plates and the knives on the right. We can suppose that in doing so, she asserts a non-moral-non-aesthetic normative proposition of kind (1b), which is made true only by the informal social convention governing table-setting.

Some games are governed by informal social conventions: hide-and-seek, for example. When the child who is *IT* starts counting, Alice might assert to a newcomer-child that he ought to hide. If she does, then she is likely to be asserting a non-moral-non-esthetic proposition that is of kind (1b).

Other games are governed by formal social arrangements—tennis, golf, and chess, for example. So when a chess teacher asserts to a student that the student ought not move his or her rook on a diagonal, the teacher is typically asserting a non-moral-non-aesthetic proposition of kind (1a).

What of non-moral-non-esthetic propositions of kind (2)? Suppose Alfred tells Alice that he wants to buy a computer, and she asserts that he ought to go to Staples and does so because she thinks he would best satisfy his want by going there. Then the proposition she asserts is arguably non-moral and non-aesthetic, and is not made true by any formal social arrangement or informal social convention. The proposition she asserts is arguably made true (if it is true) by the fact that Alfred would best satisfy his wants by going to Staples.

Again, suppose it has been raining, and Alfred gazes up at the sky, which is now clearing, and says the words "The sun ought to come out soon." He might, of course, be asserting a moral proposition to the effect that the sun is under a duty to come out soon; however he is far more likely to be asserting an epistemic proposition to the effect that it is likely that the sun will come out soon. (If he is, is the proposition he asserts normative? On some views, it is analyzable into the proposition that people in possession of the evidence he has in hand ought to believe that the sun will come out soon, and if so, then it *is* normative.)

Another kind of epistemic proposition is plainly normative, for example, the proposition (asserted by Alfred) that Bert ought not jump to conclusions, and the proposition (asserted by Carol) that Dora ought to be rational. (But are those propositions non-moral and non-aesthetic? There is room for argument that they are both epistemic and moral.)

Still other propositions asserted by people who say words of the form "A ought to φ" are about things that (like the sun) aren't human beings, yet they aren't (anyway aren't merely) epistemic propositions. The manufacturers of a brand of toaster may advertise that their toasters ought to provide many years of safe, even toasting. If they assert this, then they aren't (anyway aren't merely) asserting that their toasters are likely to provide many years of safe, even toasting; they are asserting that their toasters are required to do so. (That is why they invite you to return the toaster you buy if it turns out to be defective.) So the proposition they assert is normative, but it is neither moral nor aesthetic.

5. The existence of many kinds of non-moral-non-aesthetic normative propositions that are asserted by use of the word *ought* has tempted some philosophers to think it is multiply ambiguous: that it has a moral meaning and an aesthetic meaning, and also an etiquette meaning, a game meaning, a practical-reason meaning, one or more epistemic meanings, and a function meaning. A growing literature by philosophers and linguists argues that there are better ways to accommodate this multiplicity of uses of *ought* than by appeal to its ambiguity.

6. It should also be noticed, moreover, that *ought* is not the only word that can be used to assert moral, aesthetic, and non-moral-non-aesthetic normative propositions: the same holds of *should, must,* and *may. Good* is also used to assert moral, aesthetic, and non-moral-non-aesthetic normative propositions, for while Alfred may say the words "Bert is good" because he thinks Bert is a good person, Alice may say them because she thinks Bert dances beautifully, and Arthur may say them because he thinks Bert is good at playing chess. The same holds of *right, wrong, correct,* and *incorrect*. Charles may say the words "That was right" because he thinks that what he refers to by *That* (perhaps David's killing of Dora) was morally right, Carol because she thinks that what she refers to by *That* (perhaps the color

green) was the right color to use on David's walls, and Carl because he thinks that what he refers to by *That* (perhaps the word *Yes*) was the right answer to the question "Is David older than Donald?" If so, then they all assert evaluatives, Charles's moral, Carol's aesthetic, and Carl's non-moral and non-aesthetic.

It should be noticed too that if Carl says those words for that reason, then on any plausible view of the matter, he asserts a proposition that has a truth-value. (It is on any view true or false that *Yes* is the right answer to the question "Is David older than Donald?") If we were right to say that the proposition Carl asserts—like those asserted by Charles and Carol—is normative, then we are entitled to conclude that the expressivist is mistaken in thinking that there are no normative propositions.

Still other words—such as *able, careful, intelligent, sensible, suitable, satisfactory, defective,* and *comfortable*—are typically used to assert evaluatives that are non-moral and non-aesthetic; and many (arguably all) of those evaluatives can very plausibly be thought to have truth-values.

Our communication with each other is rich in normativity. Focusing on the moral and aesthetic alone was short-sighted: no theory of normativity is adequate unless it explains what normative propositions generally have in common.

Judith Jarvis Thomson

See also Conventions, Logic of; Law, Social Phenomenon of; Moral Cognitivism; Normativism Versus Realism; Social Conventions; Social Norms

Further Readings

Blackburn, S. (1993). *Essays in quasi-realism.* Oxford, England: Oxford University Press.

Geach, P. (1965). Assertion. *Philosophical Review, 74,* 449–465.

Gibbard, A. (1990). *Wise choices, apt feelings.* Cambridge, MA: Harvard University Press.

Gilbert, M. (1989). *On social facts.* New York, NY: Routledge.

Hart, H. L. A. (1961). *The concept of law.* Oxford, England: Oxford University Press.

Korsgaard, C. (1996). *The sources of normativity.* New York, NY: Cambridge University Press.

Thomson, J. J. (2008). *Normativity.* Chicago, IL: Open Court.

Wedgwood, R. (2007). *The nature of normativity.* Oxford, England: Oxford University Press.

OBJECTIVITY

This entry discusses the nature and place of objectivity in science, and especially the social sciences, and introduces the various issues making up the topic of objectivity as well as various critical approaches to it.

Being objective is an attribute that can be related to truth, knowledge claims, and persons. Aiming at objectivity is traditionally considered as a necessary condition of all science, but this is challenged by relativists and social constructivists. The formulation of the ideal of objectivity by scientific realists usually relies on assumptions that are typical to natural science, but it can be defended also in the case of the social sciences.

Let us briefly see what objectivity involves for the three items mentioned above: (1) truth, (2) knowledge, and (3) persons.

According to the realist theory of truth, a belief or statement is true if its content corresponds to facts obtaining in reality. A modern explication of this correspondence theory, or reality theory, of truth is given by Alfred Tarski's semantic definition: Truths have to be expressed in some interpreted language or conceptual system, but in the objective or absolute sense, truth is not relative to person, gender, social class, culture, belief, theory, or paradigm.

Knowledge, classically defined as justified true belief, is objective if it is true in the absolute sense. Most scientific realists are fallibilists, who acknowledge that all knowledge claims about matters of fact are corrigible by further inquiry. Even though complete certainty cannot be reached, the most reliable way of finding well-grounded knowledge is justification by publicly accessible scientific evidence.

According to the traditional ideal, a scientist should be an objective person: just, impartial, unbiased, not misled by emotions, personal prejudices, or wishful thinking. Thus, science is objective in so far as its results correspond to the real properties and lawlike relations of the research object, and in contrast, science is subjective insofar as its results are influenced and biased by personal motives, values, wishes, beliefs, and interests.

Objectivity and Science

Science is the systematic and institutionalized pursuit of new knowledge by using critical methods of inquiry. The object of research is a domain of reality consisting of some natural, mental, or social phenomena. Researchers propose hypotheses about this domain and test them against evidence obtained through interaction between the researchers and the studied objects. Such interactions include observations, measurements, experiments, and interviews. To guarantee that reality itself is the standard of scientific knowledge, the results of such interactions should be intersubjectively recognizable: Research data should be openly accessible; observations, public; and experiments, repeatable. The results of research constitute scientific knowledge only when they have passed the internal quality control and critical discussion within the scientific community. In this sense, the real subjects of scientific knowledge are communities of investigators rather than individual scientists.

The ideal of objectivity relies on the conceptual distinction between subject and object, but its actual realization depends on subject–object interactions. This distinction may be problematic in some natural sciences: In microphysical experiments, the measuring apparatus disturbs the measured objects by changing their properties in an unpredictable manner. This kind of disturbing effect is quite common in the human and social sciences, since people may behave in a nontypical manner in interviews or controlled experimental situations.

Scientific realism accepts the ontological assumption that nature is independent of human mind, language, and culture: Natural reality does not change when we conceptualize and describe it by means of scientific theories. In contrast, relativists contend that each theory defines its own ontology, and social constructivists have claimed that theoretical entities described by scientific theories do not exist independently of the scientists but are outcomes of "negotiations" in research laboratories.

On the other hand, ever since the time of Francis Bacon in the early 17th century, scientists have known that knowledge about causal laws allows us to manipulate and control nature. Knowledge about such causal laws is sought in applied natural sciences such as engineering sciences and environmental sciences, and the results of these sciences are applied to transform nature by various kinds of science-based technologies.

Objectivity and the Social Sciences

The ontological situation in the social sciences is more complex. Social reality is constituted by meaningful intentional actions of social agents who have beliefs, wants, and wishes. Social reality also includes institutions established and sustained by collective beliefs and interactions between individuals, and such institutions and organizations may act as social agents with their collective intentions. So the social reality is not mind independent in the same way as nature. Still, Émile Durkheim argued that social facts ought to be treated as "things" by the social sciences, as they are "endowed with a power of coercion" to control individuals. Karl Popper called the human-made reality studied by the social sciences World 3, in contrast to the physical World 1 and the subjective mental World 2.

In spite of this ontological difference, the methodological goal of objectivity can be defended in the social sciences as well. Even though social facts are complex entities involving collectives of people with their beliefs and intentions, statements about them can have objective truth values (true or false) in the correspondence-theoretical sense. For example, a statement such as "Snow is white" has relatively simple truth conditions referring to the color of snow, but the truth of a statement about the value of the euro in relation to the U.S. dollar depends in a complex way on the attitudes and behavior of the economic market. Similarly, a statement about the legal norms valid in a society refers to the practices and consensus within the relevant legal community. Furthermore, there are social laws that are usually based on temporal statistical regularities in human behavior. Even though they are not unchanging "iron laws" like the law of universal gravitation in physics, one can make true or false statements about them.

It is, furthermore, important to distinguish between value neutrality and interest neutrality of science. Scientific inquiry is not interest neutral, as the selection of the topics to be studied is influenced by the interests of the researchers and the funding organizations. Scientific knowledge, when properly warranted by evidence, can then be applied to serve various kinds of practical interests, such as emancipation from oppressive social conditions (as Jürgen Habermas has argued), but these interests cannot function as criteria in the assessment of the evidential support of the best theories. Indeed, science is valuable just because its hypotheses are accepted in a value-free way by appealing to its own "epistemic utilities" (e.g., truth, information, truthlikeness, and explanatory and predictive power, as Ilkka Niiniluoto has argued). Of course, this demand of objectivity may sometimes fail. Sociologists of science have given historical case studies to show that political interests have in fact influenced the conclusions of scientists. In its guide of medical research funding, the National Institute of Health wishes to eliminate "conflicts of financial interest." But if a bias is detected, it should be corrected by the self-criticism of the scientific community.

A famous defense of objective social science was given by Max Weber in 1904. Weber accepted the fact/value distinction and the demand of value-free social science. Ultimate values cannot be proved scientifically, so that they do not belong to the goals

or results of scientific inquiry. On the other hand, statements about instrumental value, or the relations between given ends and the rational means of establishing them, can be defended by scientific investigations. Such means–ends relations are studied in applied social sciences, like social policy studies, social work, business economics, military studies, peace research, and urban studies. Their typical results can be formulated as conditional recommendations with a value antecedent: "If you want *A* and believe that you are in situation *B,* you should do *X*." Even scientists, who do not endorse the value *A,* can defend the truth of this kind of conditional.

Society may change with the development of social science. Phenomena known as "self-fulfilling prophecies" and "self-refuting predictions" show that the publication of the results of social sciences can as a by-product change the attitudes and behavior of social agents. Applied social sciences, or "sciences of the artificial" in Herbert Simon's sense, develop systematically the opportunity of intentionally transforming society so that desired goals are effectively reached and some social problems are thereby solved.

Unlike purely natural phenomena, intentional social actions are laden with meanings. To describe and understand such actions, one has to reconstruct their meaning by referring to the concepts, beliefs, and intentions of the agents. It is sometimes argued that such ascription of meaning is always an evaluative task. So the hermeneutics of Hans-Georg Gadamer argues that the understanding of a cultural object (e.g., a work of art, historical text, or event) always involves the uncovering of its significance for our own lives. However, a distinction should be made between the actor's and the researcher's meaning. When we ask why Germany attacked the Soviet Union in 1941, we should consider the relevant political situation, including the intentions and beliefs of the leaders of the Third Reich, and hypotheses about this historical situation have a truth value independently of us and can be tested by appropriate evidence.

Special problems for the social sciences arise from the possibility that the researchers may themselves be part of the research domain under study. In many cases, there is a temporal or cultural distance between the subject and the object of research, such as when a political historian studies the French Revolution or a Protestant social anthropologist studies religions in Africa. Sometimes it is claimed to be an advantage if the researcher knows the studied phenomenon from inside, so that Buddhism should be studied by Buddhists and ethnic or sexual minorities by representatives of the same group. But such an involvement or "positioning" has also been seen to impair the impartiality of the researcher. Sandra Harding's feminist "standpoint theory" suggests that oppressed and marginalized groups at the bottom of social hierarchies have an epistemologically privileged position in grounding reliable knowledge claims, so that "we all—men as well as women—should prefer women's experiences to men's" (Harding, 1987, p. 10). This is a modification of the Marxist view that the proletariat in a capitalist society has a more adequate perspective on social problems and the progress of history than the bourgeois. The alternative view to such standpoint theories is to try to maximize objectivity by ensuring that the community of investigators has persons from a different background, so that academic education is open to all irrespective of their socioeconomic position.

Ilkka Niiniluoto

See also Feminist Epistemology; Kuhn and Social Science; Kuhn on Scientific Revolutions and Incommensurability; Normativism Versus Realism; Normativity; Relativism and Social Sciences; Relativisms and Their Ontologies; Scientific Method; Social Constructivism; Social Facts; Truth, Philosophical Theories of; Value Neutrality in Science; Weber and Social Science: Methodological Precepts

Further Readings

Brodbeck, M. (Ed.). (1968). *Readings in the philosophy of the social sciences.* New York, NY: Macmillan.

Durkheim, E. (1950). *The rules of sociological method.* New York, NY: Free Press.

Gadamer, H.-G. (1989). *Truth and method.* London, England: Sheed & Ward.

Habermas, J. (1972). *Knowledge and human interests.* Boston, MA: Beacon Press.

Harding, S. (Ed.). (1987). *Feminism and methodology.* Bloomington: Indiana University Press.

Krausz, M., & Meiland, J. W. (Eds.). (1982). *Relativism: Cognitive and moral.* Notre Dame, IN: University of Notre Dame Press.

Latour, B., & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts.* London, England: Sage.

Niiniluoto, I. (1999). *Critical scientific realism.* Oxford, England: Oxford University Press.
Niiniluoto, I. (2006). World 3: A critical defence. In I. Jarvie, K. Milford, & D. Miller (Eds.), *Karl Popper: A centenary assessment* (Vol. 2, pp. 59–69). London, England: Ashgate.
Niiniluoto, I. (2007). Evaluation of theories. In T. Kuipers (Ed.), *General philosophy of science: Focal issues* (*Handbook of the philosophy of science*) (pp. 175–217). Amsterdam, Netherlands: Elsevier.
Searle, J. (1995). *The construction of social reality.* New York, NY: Free Press.
Shapin, S. (1975). Phrenological knowledge and the social structure of nineteenth-century Edinburgh. *Annals of Science, 32,* 219–243.
Simon, H. (1981). *The sciences of the artificial* (2nd ed.). Cambridge: MIT Press. (Original work published 1969)
Weber, M. (1949). *The methodology of the social sciences.* New York, NY: Free Press.

Observation and Theory-Ladenness

In the philosophy of science, observations are said to be "theory-laden" when they are affected by the theoretical presuppositions held by the investigator. The thesis of theory-ladenness is most strongly associated with the late 1950s and early 1960s work of N. R. Hanson, Thomas S. Kuhn, and Paul Feyerabend, and it was probably first put forth (at least implicitly) by the French physicist, historian, and philosopher Pierre Duhem about 50 years earlier. The issue of theory-ladenness has played a key role in the philosophy of science. Lineages of the early treatment of theory-ladenness go back to Ludwig Wittgenstein and Willard van Orme Quine's holist view of language. In the philosophy of the social sciences, theory-ladenness has been associated with epistemic relativism.

The thesis of theory-ladenness, if true, has troublesome consequences for theory testing. If there are no theory-neutral observations, then this raises doubts about whether empirical tests can truly decide between competing theories. So if theories partially determine the meaning of observation terms, two investigators holding incompatible theories will mean different things when they use the same observational vocabulary, and if theories partially determine "what we see," two investigators holding incompatible theories will see the objects relevant for discriminating between their theories differently.

Fundamentals

Although often run together, at least two forms of theory-ladenness should be kept separate: (1) the *meaning* of observational terms is partially determined by theoretical presuppositions and (2) the theories held by the investigator, at a very basic cognitive level, impinge on the perceptions of the investigator. The former may be referred to as *semantic* and the latter as *perceptual* theory-ladenness.

A thesis that also goes under the heading of theory-ladenness may be (more appropriately) referred to as *theory dependence of instruments,* on which much discussion has focused: The investigator's confidence in the truthfulness of the results obtained with certain instruments depends on his or her having sound theories of how these instruments work. Such theories are also referred to as "background" theories. The theory dependence of instruments is particularly problematic when the background theories are the very theories that the investigator seeks to test, for in those scenarios the testing procedure is rendered circular.

Theory-ladenness should not be confused with certain other ideas. Theory-ladenness does not imply that our perceptions are *fully* determined by our theories; it does not imply that we see "*whatever* we want to see." No philosopher of science of some standing has defended such an extreme position. We cannot see flying pigs even if we had theories that told us that there were such things. On the other hand, theory-ladenness does not simply amount to perceptions being *interpreted* differently by different people. Nor is theory-ladenness the mere theoretical *guidance* of empirical inquiries—that is, the decision to perform certain experiments rather than others or to investigate a certain aspect of the world. Both of these ideas are platitudes and philosophically not particularly interesting.

A gray area is the phenomenon of negative theoretical bias—in other words, the idea that empirical results not amenable to certain theoretical presuppositions are (wilfully or subconsciously) ignored by the investigator. Clearly, also in cases of theoretical bias, theoretical presuppositions impinge on the

data in ways that are comparable with the thesis of theory-ladenness. Yet negative theoretical bias is normally taken to be easily revealed through various control mechanisms in scientific practice (e.g., peer review). Since theoretical bias as a form of theory-ladenness has received rather little attention by philosophers of science, it will not be discussed here.

Semantic Theory-Ladenness

One of the best-known examples for semantic theory-ladenness concerns the (observational) term *mass,* which has a different meaning in Newtonian physics from what it does in Einstein's theory of relativity. Whereas in the former theory, mass is a constant, in the latter it depends on the velocity of the object in question. Or take the term *planet*. In the Ptolemaic system, the (observational) term *planet* referred to a class of astronomical objects that included the sun and the moon but not the earth. In contrast, in the Copernican system, the term *planet* included the earth but neither the sun nor the moon. Kuhn used examples like these to argue for his (perhaps most) controversial idea of the incommensurability of paradigms. Kuhn's view has been rejected by many philosophers on the grounds that he employs a particular theory of reference (descriptivism/holism) that most philosophers take to be inadequate. Still, largely due to Kuhn, Hanson, and Feyerabend, the vast majority of philosophers have accepted that there are hardly any observational terms relevant to scientific practice that are not theory-laden. Indeed, this was one of the major reasons that led to the demise of logical positivism/empiricism, which had postulated a strict distinction between observational and theoretical vocabulary, whereby any theoretical terms had to be relatable to observational terms for the former to be deemed meaningful. Modern-day empiricists such as Bas van Fraassen, however, do accept semantic theory-ladenness, the vagueness of the observation/theoretical distinction, and that theoretical terms are meaningful without being relatable to observational terms, but they do insist on skepticism about the unobservable referents of theoretical terms (e.g., "electron").

Perceptual Theory-Ladenness

An instructive example for perceptual theory-ladenness is a psychological experiment famously used by Kuhn (1996) in his *Structure of Scientific Revolutions*. In this experiment, the subjects were presented with a set of playing cards (one by one) that contained "anomalous" cards, such as the black four of hearts. Interestingly, with a sufficiently short amount of exposure, the subjects perceptually assimilated the anomalous playing cards to the normal ones; that is, they would report a black four of spades when being presented with a black four of hearts. Apparently, their belief that playing cards fall into certain categories "primed" their perceptual system accordingly. Other examples illustrating the second form of theory-ladenness include "Gestalt figures" like the duck/rabbit or the Necker cube, which are drawings that can be perceived in two different ways without the object of perception changing. These examples are, of course, merely illustrative. An example from scientific practice allegedly exhibiting Gestalt-like features, given by Hanson and Kuhn, is the difference between a Copernican (a believer in a sun-centered universe) and a Ptolemaian (a believer in an earth-centered universe) seeing different things when looking at a sunset. Whereas the Ptolemaian sees the sun falling behind the horizon (because in his or her view, the sun is moving, not the earth), the Copernican sees a fixed sun and a rising horizon. Likewise, an Aristotelian, who believes that all objects have the natural tendency to fall to the earth, sees the "constrained" fall of an object when watching a pendulum, whereas a Galilean, having developed an early form of the concept of inertia, sees damped inertial motion.

A straightforward criticism that can be leveled against those latter examples in particular is that they are based on a confusion between "seeing" and "seeing *that,*" the latter being a propositional attitude that requires judgment. Bas van Fraassen, in his influential *The Scientific Image,* illustrates this point with the example of an aboriginal tribe seeing a tennis ball for the first time: Although they do, as we would, see the tennis ball (i.e., a yellow fluffy thing that bounces), they don't see *that* this is a tennis ball, for this would require some basic familiarity with the game of tennis, immersion in Western culture, and so on. Likewise, Antoine Lavoisier and Joseph Priestley, as much as Aristotle and Galileo (both persons in each pair), would see the same objects, but each of them would make different (theoretically informed) judgments about them. However, there are examples where this interpretation seems less

plausible. A native speaker of Chinese, for instance, just *hears* (without needing to interpret or consciously judge) meaningful utterances when hearing Chinese rather than hearing a mere succession of vocal sounds, as a speaker without any knowledge of the Chinese language would. In examples like these, theory-laden observation has a character of immediacy and inevitability that is not reflected by the "seeing" versus "seeing that" distinction.

Another related critique of perceptual theory-ladenness was given by Jerry Fodor in the 1980s. Based on examples such as the Müller-Lyer illusion, in which two lines of equal length appear to be of different length, Fodor argued that perceptions are *cognitively impenetrable*. That is, even if we are made aware of the two lines being of equal length, we still perceive them as being of unequal length. Fodor's attack on perceptual theory-ladenness has been taken up by others in more recent works. Two things should be noted about Fodor's defense. First, appeal to examples such as the Müller-Lyer illusion can only show that *some* perceptions are cognitively impenetrable. Second, as indicated above, defenders of perceptual theory-ladenness do not claim that all of our perceptual experience is subject to cognitive penetration. If that were so, it would indeed be possible that we see what we wish to see. Again, no philosopher of science has defended such a view. On the contrary, the Kuhnian account of science, for instance, presupposes that there are observations that *resist* theoretical assimilation ("anomalies").

Still another influential way of countering perceptual theory-ladenness was made by Bogen and Woodward (also in the 1980s). They advance the view that scientific theories explain and predict unobservable phenomena rather than observable data. Phenomena are inferred from data (usually by statistical methods). But if this is so, neither phenomena nor data can be theory-laden: Phenomena are not perceptually theory-laden because they are unobservable, and data are not theory-laden because they do not form the basis against which we test theories. Bogen and Woodward's simplest example for an unobservable phenomenon is the melting point of lead, at 327.5 °C, which is inferred from individual data points, none of which might exhibit this exact value. A more advanced example for a phenomenon is the weak neutral current, which was inferred from bubble chamber pictures (the data). The Glashow-Salam-Weinberg model explains and predicts the neutral current but not the bubble chamber pictures.

Theory Dependence of Instruments

Paul Feyerabend, in his *Against Method,* pointed out that Galileo, when gathering telescopic observations in support of the sun-centered universe in the early 17th century, had no knowledge of the working of the telescope. Among other things, Feyerabend reasons that Galileo would have needed such a theory to provide good grounds to convince his skeptical contemporaries of the truth of heliocentrism. Pointing the telescope to terrestrial objects to demonstrate its magnifying effect would not have sufficed since his contemporaries thought that the physics on earth was completely different form the physics of the heavens. In fact, Feyerabend accuses Galileo of circular reasoning: For Galileo's telescopic observations to be acceptable evidence for his contemporaries, Galileo had to show the inadequacy of the Aristotelian "two-physics" world picture. But to show this, he relied on his telescopic observations, which, again, presupposed that the physics on earth and the physics of the heavens were the same.

Several strategies have been proposed to counter the alleged theory dependence of instruments leading to the circularity of testing procedures. Some philosophers, such as Peter Kosso, have accepted the theory dependence of instruments and have demanded that the tested theory be *independent* from the theories that are presupposed when instruments are used. Others have simply denied that one needs knowledge of how an instrument works to *use* that instrument in such a way that it produces reliable data. Ian Hacking's *Representing and Intervening* has been influential in this respect. He points out that observations made with microscopes in the 19th century were trusted despite no accurate theory of the microscope being available. Hacking also argues that the observations made by biologists when using microscopes are not undermined by the fact that biologists regularly know rather little of the physics that underlies these microscopes. And yet *someone* knows about it. But even if the reliability of experimental results depended on background theories, Hacking argues, scientists have a powerful strategy for addressing this problem. This

is Hacking's famous "argument from coincidence," which is also known as the "robustness" argument. It says that it would be a preposterous coincidence if the observations made with *several* instruments, each presupposing *different* background theories, were to converge and the observations were not reliable. Hacking's example is the presence of red blood platelets in the same locations (of a grid) when viewed with the optical and the electron microscopes. The robustness argument has been complemented with various other "epistemological strategies" in the work of Allan Franklin.

Still other philosophers have not been convinced that theory dependence of instruments of the circular variety should be taken seriously. Hasok Chang in his *Inventing Temperature* considers the example of establishing that mercury expands uniformly with a rise in temperature for the production of reliable thermometers in the 19th century. To do this, one had to plot the volume of mercury versus temperature. But for that one, of course, needs a reliable thermometer, which was what scientists had set out to discover in the first place. But Chang argues that this kind of circularity is innocuous: Even if the theory assumed when using a particular instrument is the same as the one at stake, the actual experimental result is still contingent. In other words, the reliance on assumptions that we wish to establish when using a certain instrument does not imply that the results gained with those instruments are guaranteed to be of a particular form.

Samuel Schindler

See also Duhem-Quine Thesis and the Social Sciences; Experiment, Philosophy of; Feyerabend, Critique of Rationality in Science; Holism, in the Philosophy of Language; Instrumentalism of Scientific Theories and Constructive Empiricism; Kuhn on Scientific Revolutions and Incommensurability; Logical Positivism/Logical Empiricism; Relativism in Scientific Theories; Scientific Method

Further Readings

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *Philosophical Review, 97*(3), 303–352.

Chang, H. (2004). *Inventing temperature: Measurement and scientific progress:* New York, NY: Oxford University Press.

Feyerabend, P. (1975). *Against method*. London, England: Verso.

Fodor, J. (1984). Observation reconsidered. *Philosophy of Science, 51,* 23–43.

Franklin, A. (2008). Experiment in physics. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy* (Spring 2010 ed.). Stanford, CA: Stanford University, Center for the Study of Language and Information, Metaphysics Research Lab. Retrieved from http://plato.stanford.edu/archives/spr2010/entries/physics-experiment

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science* (Vol. 355). Cambridge, England: Cambridge University Press.

Hanson, N. R. (1958). *Patterns of discovery: An inquiry into the conceptual foundations of science*. Cambridge, England: Cambridge University Press.

Kosso, P. (1989). Science and objectivity. *Journal of Philosophy, 86*(5), 245–257.

Kuhn, T. S. (1996). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Machamer, P. K. (1973). Feyerabend and Galileo: The interaction of theories, and the reinterpretation of experience. *Studies in History and Philosophy of Science, 4,* 1–46.

Schindler, S. (in press). Theory-driven data reliability judgments: Theory-ladenness in the experimental context?

Schindler, S. (in press). Theory-laden experimentation. *Studies in History and Philosophy of Science*.

Stegenga, J. (2009). Robustness, discordance, and relevance. *Philosophy of Science, 76*(5), 650–661.

Suppe, F. (1977). *The structure of scientific theories*. Champaign: University of Illinois Press.

Oppression

The concept of oppression has been controversial but powerful in the history of political philosophy and is characterized differently by Marxists, liberals, feminists, and race theorists. As it is used today, "oppression" refers to a harm in which groups of persons are systematically and unfairly or unjustly constrained, burdened, or reduced by social forces. The concept is thus fundamentally normative but makes important descriptive claims that engage social-scientific explanations. The main controversies about oppression include how oppressive institutions arise, which

groups are oppressed and why those groups, the main causes and harms of oppression, and the prospects for ending oppression.

The philosophy of oppression begins in the modern period with liberalism, which recognized the rights of individuals to freedom and justice. To modern liberal political theorists, "domination," "tyranny," and "oppression" were synonyms, connoting rule by an arbitrary or opposing will, resulting in violation of liberal political rights, economic deprivations, and physical brutality. They conceived oppression in individualistic terms, and its harms included being ruled by illegitimate governments and religious intolerance by the state. In the 19th century, the scope of liberal political rights broadened in the work of G. W. F. Hegel and Karl Marx and by works on women's rights and abolitionism. These theories posed a more social conception of oppression, where oppressor and oppressed may be related in a less politically formalized way. For Hegel, oppression was a failure to recognize the equal moral worth and dignity of another. Marxist, abolitionist, and feminist writers described oppressions of one social group by another (nongovernmental) social group.

Marx's historical materialism is the first attempt to scientifically explain the origin and maintenance of oppression. Marx understands oppression as causally based in the economic system of the epoch. Exploitation, a form of oppression, begins with division of labor and thus with the ability of one group to coercively appropriate the product of another's labor. In capitalism, the working class is systematically and materially disadvantaged through the organization of production, in which the capitalist class appropriates the surplus value of workers.

Contemporary discussions of oppression extend this analysis and provide different causal accounts of its origin and endurance in different oppressed groups. Three fairly standard conditions characterizing oppression can be summarized as follows:

1. *The harm condition:* A wrongful harm that comes out of an institutional practice
2. *The coercion condition:* Unjustified coercion or force that brings about the harm
3. *The social group condition:* Harm that is perpetrated on a social group

Condition 3 may seem to posit social collectives that exist over and above the individuals that make them up, which would raise ontological worries. Philosophers have shown how social groups can be constructed from individuals and the causes and effects of their actions.

Many accounts of oppression point out that when one social group is oppressed, another is typically raised up—materially, culturally, or psychologically. This extra boost to the nonoppressed has come to be known as "privilege" and can be stated as a fourth defining condition:

4. *The privilege condition:* Another social group that benefits from the institutional practice in Condition (1)

Recent theories have explored the significant psychological causes and effects of oppression. Some theorists of oppression use psychoanalysis to explain how individuals in certain social groups are motivated to be dominant or submissive, particularly gender groups and colonized national groups. Some contemporary thinkers theorize oppression as a problem arising from misrecognition, dehumanization, or failure of cultural respect. These theories share the core idea with Hegel that to be oppressed is fundamentally to be denied dignity or equal moral worth, either as individuals belonging to a social group or as a group as a whole.

Other theories are more concerned about the economic inequalities of oppression. Women's oppression on these theories is both an economic constraint and a psychologically degrading and distorting force. Theories of racial oppression have focused on dehumanization and invisibility as harms. Finally, Marxist and liberal political theorists, insofar as the latter treat aspects of oppression at all, tend to give primacy to the problems of economic distribution.

Another important development in recent theories of oppression is the recognition that persons belong to intersecting social groups, some of which may be oppressed while others may be relatively privileged. For example, Black men are both oppressed as Blacks and privileged as men, although intersecting group status does not always combine in the same way, and some privilege can turn to disadvantage under certain circumstances. Black feminist theorists originated this so-called intersectional analysis of oppression, which has now become standard, especially in feminist discussions of oppression and in discussions of the nature of race and ethnicity.

A controversy about oppression concerns its inevitability in human society. According to one prominent evolutionary psychological account of social dominance, there will always be a dominant and a dominated social group as a result of the basic structure of human psychology. Human males have evolved to form in-groups in order to have exclusive sexual access to the females in their in-group, the boundaries of which they will then violently enforce. The theory explains all forms of social group domination as a matter of evolved tendencies of males to police in-group/out-group boundaries. Since females have fewer tendencies toward domination, the theory predicts that they are easily dominated by males and that females are less likely to dominate others of any group. The evolutionary approach to explaining oppression equates oppression with dominance, making it impossible to distinguish justified from unjustified forms.

Finally, there are libertarian concerns about the claim that there are oppressed social groups, as opposed to simply individuals who are unjustly treated. However, recent work on stereotype threat and implicit bias, and a well-developed literature on stereotyping by psychologists show that individuals are perceived by others as belonging to social groups and are treated in stereotypical and biased ways as a result. There can be little doubt that individuals' social group status affects their opportunities in ways that are completely out of their control. Thus, social group status must be seen as an important locus of oppression.

Ann E. Cudd

See also Evolutionary Psychology; Feminism: Schools of Thought; Group Identity; Implicit Bias and Social Cognition; Marxism and Social/Historical Explanation; Prejudice and Stereotyping; Race, Theories of

Further Readings

Cudd, A. E. (2006). *Analyzing oppression*. New York, NY: Oxford University Press.

Davis, A. Y. (1981). *Women, race, and class*. New York, NY: Random House.

Fanon, F. (1963). *The wretched of the earth*. New York, NY: Grove Press.

Marx, K. (1978). *The Marx-Engels reader* (2nd ed.; R. Tucker, Ed.). New York, NY: W. W. Norton.

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York, NY: Cambridge University Press.

Paradigms of Social Science

Thomas Kuhn's *The Structure of Scientific Revolutions*, published in 1962, introduced the word *paradigm* to discussions of the nature of science. The term rapidly passed into wide circulation, accompanied by complaints that Kuhn in fact used it promiscuously and confusingly. Kuhn accepted some of that criticism and responded that he had primarily intended to identify two main features of mature natural sciences, which could be differentiated into (1) a disciplinary matrix and (2) an exemplar. Both were meant to identify a way in which mature natural sciences, *unlike social studies such as sociology*, possessed a large measure of agreement on their fundamentals, which could be taken for granted in developing and conducting investigations into phenomena. *Disciplinary matrix* was meant to index a substantial measure of agreement on the aims and purposes of their field among the scientists working in it—agreements on central concepts and methods of investigation. *Exemplar* emphasized the extent to which a developed branch of natural science is unified around a quite specific idea of how research is to be done, which was defined by a distinguished and potent achievement—hence "exemplar"—that was accepted as a model for how research was to be done.

Kuhn's work was always controversial and led, quickly, to questions about whether the social sciences were as different from the natural sciences as Kuhn had made them seem. If one understood the question "Does sociology have paradigms?" in the sense of "Does it have a disciplinary matrix?" then the ready answer would seem to be yes. There are shared positions within sociology, ones that are often called "schools" or "approaches," many of which could as well be called disciplinary matrices. Then, one could talk about sociology as a multiparadigm science, having in mind a wide assortment of positions including Marxism, functionalism, rational action theory, interactionism, structuralism, and postmodernism, among others. These are marked by some internal agreement on broad and significant matters. However, it should not be assumed that the agreement within these "disciplinary matrices" is especially deep and strong, and in any case, there is reciprocal mutual skepticism between these positions. In relation to Kuhn's revised terminology, it would seem that a social science such as sociology can, like a natural science, have paradigms.

Kuhn's initial intent was to contrast natural with social science, and calling sociology a multiparadigm discipline does not mitigate the contrast. Kuhn's point about *mature* branches of natural science was that they were internally unified and would typically have only one paradigm—division into disputing paradigms was only an occasional and transient feature in times of "Scientific Revolution." The usual and defining condition was that of "normal science."

The weight of Kuhn's contrast, in terms of both the disciplinary matrix and the exemplar, was to emphasize how close the agreement was among researching scientists, so close that scientists did not have to spend large parts of their time arguing

about and trying to justify what they took to be the fundamentals of their discipline. Nor did they have to work out for themselves how to implement their general ideas and methods, for the exemplars embodied quite concrete and precise displays of the forms of investigation currently acceptable in the field and thereby solved such problems. The exemplar thus plays the key part in enabling normal science, a period in which researchers can take close agreement with their colleagues for granted and can apply accepted standard methods in the investigation and solution of empirical problems, confident that success in applying those procedures will ensure that their results are recognized as additions to their area's knowledge.

A multiparadigm discipline like sociology is not like that, as sociologists very well know. The sociological paradigms nominated above do not embody exemplars, and many sociologists complain bitterly over sociology's inability to develop an accumulating body of empirical knowledge (rather than just a heterogeneous assortment of largely unrelated studies), to enter, in Kuhnian terms, a period of normal science.

Is there anything in sociology that at least approximates Kuhn's formal criteria for exemplars? Some good candidates tend to be rather marginal to sociology's mainstream, as for example, conversation analysis (CA, the study of talk-in-interaction), which is perhaps more influential outside sociology, in fields like linguistics and psychology, than it is in sociology, and expectation states theory (EST, the study of formal properties of hierarchical relations), which is largely disregarded outside its own boundaries. In CA's case, the article "A Simplest Systematics for the Organization of Turn Taking in Conversation" by Harvey Sacks, Emanuel Schegloff, and Gail Jefferson (1974; first published in a linguistics journal) and Sacks's collected *Lectures on Conversation* (1986) provide a model for analyzing the way the distribution of turns at talk in conversation and other "speech exchange systems" (such as courtroom questioning or therapeutic dialogues) is organized as well as how turn distribution shapes innumerable aspects of what is done in those turns. This model has been closely followed in a large and increasing number of studies by other researchers.

EST, originating in experimental studies of status relations in small groups by Joseph Berger in the 1950s, has tried to develop a systematic and formal theory together with an integrated set of studies to show how individuals' expectations based on assessments of others' status affects group formation and structure. Much of that project's work is recorded in volumes of *Sociological Theories in Progress*. There undoubtedly are other areas of sociology, such as the more prominent *social network theory*, that could also qualify. How such cases compare with the natural science instances Kuhn had in mind and whether they do involve genuine accumulations of knowledge are questions that cannot be explored here; but even if they do, these cases are not typical representatives of sociology's multiparadigm situation.

Wes Sharrock

See also Kuhn and Social Science; Kuhn on Scientific Revolutions and Incommensurability; Philosophy of Sociology, History of; Reductionism in the Social Sciences; Sociology of Knowledge and Science

Further Readings

Berger, J., Zelditch, M., & Anderson, B. (Eds.). (1966). *Sociological theories in progress* (Vol. 1). New York, NY: Houghton Mifflin.

Berger, J., Zelditch, M., & Anderson, B. (Eds.). (1972). *Sociological theories in progress* (Vol. 2). New York, NY: Houghton Mifflin.

Berger, J., Zelditch, M., & Anderson, B. (Eds.). (1989). *Sociological theories in progress: New formulations*. Thousand Oaks, CA: Sage.

Bryant, C. (1975). Kuhn, paradigms and sociology. *British Journal of Sociology, 26*(3), 354–359.

Carrington, P., Scott, J., & Wasserman, S. (Eds.). (2005). *Models and methods in social network analysis*. Cambridge, England: Cambridge University Press.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Masterman, M. (1970). The nature of a paradigm. In I. Lataktos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 59–90). Cambridge, England: Cambridge University Press.

Sacks, H. (1986). *Lectures on conversation* (2 vols.). Oxford, England: Blackwell.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language, 50*(4), 696–735.

Sharrock, W., & Read, R. (2002). *Kuhn: Philosopher of scientific revolutions*. New York, NY: Wiley.

Pareto Optimality

This entry explains the notion of Pareto optimality (after Vilfredo Pareto [1848–1923], the Italian economist, sociologist, and engineer), the normativity of Pareto-optimal outcomes, the function of economic actors' revealed preferences, and the role of Pareto optimality in welfare economics. The entry ends by looking at two main criticisms.

Pareto optimality (Pareto efficiency, allocative efficiency) is an important concept in economics and related social sciences, where it figures in both normative and descriptive claims. To understand Pareto optimality, it helps to understand *Pareto improvement* (PI). A situation *A* is a PI over a situation *B* if and only if at least one person (strictly) prefers *A* to *B* and no one (strictly) prefers *B* to *A*. A situation is Pareto optimal just in case no other situation is a PI over it—that is, no one can get more preference satisfaction without someone else getting less. Note that multiple situations can be Pareto optimal.

Desirability of Pareto Optima

Economists generally see Pareto optimality as desirable. This view is expressed in a number of ways—Pareto-optimal outcomes are *socially rational*; non–Pareto-optimal outcomes ought not to be chosen—but such claims are best seen as corollaries of a basic assertion about the value of PIs: If *A* is a PI over *B*, then *A* is *better* than *B*. A PI seems to make at least one person better off and no one worse off, and so to result in an uncontroversial gain in overall well-being. Human welfare has moral significance; such a gain is good, at least other things being equal. The *other-things-equal* (ceteris paribus) clause is required because well-being is not the only moral consideration: Fairness, justice, equality, and so on, also matter. If PIs are good (in this limited sense), then so are Pareto optima: Each is the end point of a series of PIs such that all of the uncontroversial preference gains have been achieved.

Preferences and Market Competition

Along with many other social scientists, economists hold that a person's preferences are revealed by her behavior. This allows them to determine the distribution of preference satisfaction, and so identify Pareto optima, by merely observing what people do. PIs are defined in terms of preferences precisely because the preference-to-behavior connection makes the existence of PIs empirically testable. Furthermore, economists think that they can identify social arrangements that lead to Pareto optima. They focus, in particular, on *perfectly competitive markets*. The basic idea behind the claim that perfect competition leads to Pareto optimality is simple: Voluntary trades are PIs; when no one is willing to trade further, the situation will be Pareto optimal. Perfect competition is defined to ensure that all voluntary trades are PI (e.g., no "entangled" preferences, so mutual agreement among traders suffices for a PI) and that no PI trades are blocked (i.e., no barriers to entry, so everyone can seek trading partners). Perfect competition is not required for Pareto optimality; central planning might suffice—for example, parental allocation of resources to children. Perfect competition stands out, however, because it has an internal dynamic that tends toward Pareto optimality: People trade for their own satisfaction, without any external incentives.

Welfare Economics

Both the normative and the descriptive aspects of Pareto optimality play a crucial role in standard welfare economics. The *fundamental theorems of welfare economics* spell out the usual approach. The first holds that the outcomes of perfect competition are Pareto optimal. The second holds that any attainable Pareto optimum can be achieved through perfect competition from an appropriate initial distribution of resources. The welfare theorems are usually taken to establish that a division of labor is appropriate in the realm of interpersonal morality: Efficient markets should handle welfare issues; lump-sum resource transfers should handle other moral considerations (e.g., fairness), without interfering with market mechanisms.

Critical Points

There are many criticisms of economic approaches where Pareto optimality plays a significant role, but this entry will discuss only two, which are perhaps the most common.

The first is that PIs are (practically) *unattainable*. To identify market transactions as PIs, economists typically assume that people care only about the trades in which they participate. But, in fact, people

are concerned with the transactions of others. Envy comes to mind, but there are also moral attitudes—many would prefer to see less self-degradation, unfairness, and so on, even when they aren't personally involved. Such entangled preferences undermine the connection between market transactions (even if unconstrained) and uncontroversial welfare gains since many voluntary trades will no longer be PIs. If most situations qualify, then being Pareto optimal is not much of an achievement.

A second criticism of Pareto-optimality claims in social science holds that there is no regular connection between preference satisfaction and welfare. The clearest illustrations involve people with implausible views about what benefits them, such as Scrooge from Dickens's *A Christmas Carol*. The distinction between satisfying preferences and actually being better off doesn't, however, depend on a particular view of the good. For any view, even someone who holds it will prefer some things that are inconsistent with that conception. In the most prosaic cases, false beliefs lead a person to prefer one thing when it would make sense (given their more basic values) for her to prefer another. Furthermore, psychology tells us that people have trouble bridging the gap between ultimate goals and situation-specific preferences; for example, they have a hard time resisting nearer but lesser goods. Reasoning goes awry in many ways, so it is practically certain that people will prefer things that are not beneficial on their own view of the good. Furthermore, there is no way to bridge the gap between the preferred and the beneficial without making the distribution of preference satisfaction irrelevant. In order to separate value-congruent from value-incongruent preferences, one must be explicit about the value standards at issue and use them to evaluate preferences. Once such standards are at hand, however, there is no point in consulting preferences—one looks directly at whether behavior advances the relevant values.

Stephen Ellis

See also Cost–Benefit Analysis; Homo Economicus; Invisible Hand Explanations; Markets and Economic Theory; Preference; Sen's Paretian Liberal; Welfare Economics

Further Readings

Hausman, D., & McPherson, M. (2006). *Economic analysis, moral philosophy, and public policy* (2nd ed.). New York, NY: Cambridge University Press.

Heyne, P. (1998, December). Moral misunderstanding and the justification of markets. *The Region*. Retrieved from http://www.minneapolisfed.org/publications_papers/pub_display.cfm?id=3583

Rescher, N. (1979). Economics vs. moral philosophy: The Pareto principle as a case study of their divergent orientation. *Theory and Decision, 10,* 169–179.

Sen, A. (1985). The moral standing of the market. In E. Paul, F. Miller, & J. Paul (Eds.), *Ethics and economics* (pp. 1–19). Oxford, England: Blackwell.

Path Dependence

The outcome of a *path-dependent* process depends on the course taken to reach that outcome and, in particular, on contingencies that have influenced the process prior to its reaching its outcome. The concept has attracted much interest among social scientists, partly because of the possible suboptimality of the outcome of a path-dependent process, a possibility that conflicts with the claims often made by economists of social processes and the "equilibria" on which they converge.

This entry examines what path dependence is, looks at a well-known case of path dependence, and considers the relationship between the concept and that of *market failure*.

Put negatively, path-dependent processes do not converge on a unique terminus irrespective of the "historical accidents" that occur on the way. This may be expressed by describing path-dependent processes as *nonergodic*. Path-dependent processes are consequently unpredictable (although not ontologically indeterminate), for without knowledge of the contingencies that occur along the path prior to their occurrence, an observer can predict neither the course of the path nor its resting place. Studying cases of path dependence therefore involves historical analysis, whereby such contingencies and their influence on the course of the process are traced ex post facto. This explains why studies of path dependence are associated with the motto "History matters": Only if a process converges uniquely to a given point (i.e., if the process is path *in*dependent) may one ignore the history of the process and the apparently "random" events that characterize its course.

The concept of path dependence has become common to studies of economic and technological

development since the 1980s, although Thorstein Veblen is usually acknowledged as recognizing the phenomenon of path dependence (though not by name) in 1915. The concept may be said to have "come of age" with the works of Brian Arthur and Paul David. The latter's account of the evolution of the QWERTY typewriter keyboard is the most famous study in which claims of path dependence are made. The QWERTY configuration was initially designed in the latter half of the 19th century to overcome a technical problem whereby the typebars became jammed during typing. The QWERTY configuration overcame this problem, yet it is technically inferior to other keyboard settings, as evinced by speed-typing world records, which have usually been claimed by typists using not QWERTY but the Dvorak keyboard, patented in the 1930s. One is thus led to ask why QWERTY is still the dominant keyboard design. One "historical accident" that provided a decisive fillip to QWERTY was the development in the 1880s of touch typing, originally done on the QWERTY keyboard. The association between QWERTY and the inception of touch typing meant that the early touch typists were QWERTY trained, and so those buying typewriters—primarily businesses rather than private households—had an incentive to install QWERTY equipment so that they could employ the most efficient (QWERTY) typists. This was obviously of advantage to QWERTY-trained typists; it also meant that those who sought a career as a typist were more likely to favor QWERTY given the increasing stock of typewriters used in business that were compatible with the skills of a QWERTY-trained typist. Every touch typist who opted for QWERTY increased the probability that others would follow suit and opt for QWERTY when making a choice between keyboard formats on which touch-typing skills were to be acquired, a selection process that has been likened to a Pólya urn process in mathematics. There was therefore positive feedback between businesses that installed QWERTY and typists who learned to touch type on QWERTY keyboards. Keyboard design thereby became "locked in" to QWERTY at the end of the 19th century, a terminus to this developmental path that has never been reversed. It is not clear how or whether convergence on the QWERTY keyboard could be reversed through a decentralized (market) process of the sort that led to QWERTY's dominance.

Further studies of path dependence have been undertaken for nuclear reactors, video recorders, and petroleum-driven cars. The concept of path dependence has also been applied to the development of social institutions, government policy, and scientific knowledge. In such studies, path dependence is often associated with inefficiency or "market failure" because decentralized market processes can fail to lead to the adoption of the optimal technology (as is claimed of QWERTY). Some economists, such as Stan Liebowitz and Stephen Margolis, have challenged such claims. One must note, though, that theorists of path dependence do not claim that a path-dependent process always or necessarily leads to an inefficient outcome, for path dependence is neither a necessary nor a sufficient condition for suboptimality. Nevertheless, paths that appear to lead to inefficient outcomes have captured the attention of social scientists, particularly of those who cast doubt on the "optimal" nature of decentralized coordination in the market and on the claims made on behalf of the market by neoclassical economic orthodoxy.

In a short space of time, *path dependence* has earned itself a place among key concepts in economics and technology studies and also, though to a lesser extent, in other social-scientific disciplines.

Mark Peacock

See also Determinism; Equilibrium in Economics and Game Theory; Feedback Mechanisms and Self-Regulatory Processes in the Social Sciences; Heterodox Economics

Further Readings

Arthur, W. B. (1994). *Increasing returns and path dependence in the economy.* Ann Arbor: University of Michigan Press.

David, P. A. (1985). Clio and the economics of QWERTY. *American Economic Review, 75,* 332–337.

David, P. A. (2001). Path dependence, its critics, and the quest for "historical economics." In P. Garrouste & S. Ioannides (Eds.), *Evolution and path dependence in economic ideas* (pp. 15–40). Cheltenham, England: Edward Elgar.

Liebowitz, S. J., & Margolis, S. E. (1995). Path dependence, lock-in, and history. *Journal of Law, Economics and Organization, 11,* 205–226.

Veblen, T. (1966). *Imperial Germany and the Industrial Revolution.* Ann Arbor: University of Michigan Press. (Original work published 1915)

Performative Theory of Institutions

This entry presents Barry Barnes's theory of social institutions, known as the *self-referential model* or sometimes as the *performative model*, and explains how it differs in important ways from similar accounts of self-reflexivity involved in constructing the social world.

In everyday language, the word *institution* refers to things such as hospitals, prisons, banks, and universities. Sociologists and anthropologists use the word in a more extended sense and include, for example, the institution of gift giving, the institution of promising, and the institution of slavery. They also include forms of knowledge and culture, such as the institution of witchcraft. The philosopher Ludwig Wittgenstein even spoke of rules as institutions.

A list of examples does not tell us what sorts of things social institutions are. If air and water are made of atoms and molecules, what are institutions made of? Are they just numbers of people who behave in regular and predictable ways, or is more involved than mere regularity and predictability? It can be strangely difficult to marshal our intuitions on these questions. Institutions involve things such as status, authority, role, and obligation, and all of these are oddly intangible. The force of an obligation, for example, the obligation to respond to a gift, is not like the force of gravity. We may say that obligations have a moral rather than a physical character, but that merely reopens the original question: What analysis is to be given to such realities?

Barry Barnes's Account

The account of social institutions and social ontology to be given here derives from the work of the contemporary sociologist Barry Barnes, though pointers toward his approach can be found in David Hume, Émile Durkheim, Ludwig Wittgenstein, and Wittgenstein's pupil and editor Elizabeth Anscombe.

Linguistic Predicates and Collective Agreement

Barnes begins by analyzing the employment of the verbal labels by which reality in any form can be identified and described. He notes that there are two different sources of information that are used when responding to an object in the environment, for example, when one wishes to label or classify it. One source is the experience of the object itself and its empirically discernible properties. The other source is the response made to the object by other people. The first method involves looking at the object; the second method involves looking at the people surrounding the object.

These two resources give rise to two different sorts of highly idealized predicate. (Barnes acknowledges that the simplifications involved here are extreme but argues that they serve to focus attention on important dimensions of concept application.) Let a word applied on the basis of the empirical properties of an object be called an N-predicate and a word applied on the basis of other people's applications of the word be called an S-predicate. Thus, I may apply the label *cat* to objects that look similar to other cats I have encountered, and I may apply the word *slave* to objects that other people call "slaves." Thus far, we can say that "cat" is an N-predicate and "slave" is an S-predicate.

In reality, the word *cat* must have been introduced by a process of ostensive definition in which a finite number of authoritatively identified cats are presented to act as models for future applications of the word. The subsequent application of "cat" must also be subject to standards of correct and incorrect use. Objective limits must be placed on the subjective intuitions of similarity guiding the application of the word and an account given of the nature of the objectivity.

To meet these requirements, some authority must (a) select certain exemplary instances of cats to provide the stereotype against which putative instances are judged and (b) other concept users must monitor the boundaries of accepted cat-hood. Both of these activities involve *collective agreement*; that is, they depend on noting and conforming to how other people respond to certain objects in the world. *Cat* is the right word to use because other people use it like this. Thus, real natural-kind terms depend on the very mechanisms that were identified as lying at the basis of S-predicates. They involve not only a response to the material environment but a response to the *social* environment. Real predicates that are meant to capture empirical features of the material world can be thought of as *combining* the machinery of pure N-predicates and that of the so-called S-predicates.

Reference

More needs to be said about the mechanics of S-predicates. What is the specific reality to which S-predicates refer? What reality corresponds to the collective decision that some objects are properly called "cats"? In what does the propriety consist? More generally, what reality corresponds to the assignments of status, such as "slave"? To call these S-predicates indicates that they are not responses to the empirical properties of things but to other people's use of the predicate. In this case, these other uses of the predicate by other predicate users must have precisely the same character; that is, they too are responses to other people's use of the predicate. There is therefore only one consistent answer to the question about the reality of the reference and the object of the response. The use of an S-predicate by one person must be a reference to the use of that predicate by everyone else. Everyone else's use has the same character, so that, taken collectively, the system of usage as a whole is a self-referential system. *The only reality to which the predicate can refer must be the reality of the system of usage itself.*

Social Reality

Here then is Barnes's general model of social reality. Social reality is created by the references to that reality, where *reference* is used broadly to cover all orientations to, and all acts invoking, that reality. An institution is a pattern of *self-referential activity*. A status is created by its being accorded, acknowledged, invoked, and acted upon. A person is an "authority" because he or she is deemed to be an authority and deferred to as an authority. A rule exists in and through its being appealed to, cited, applied, and challenged. The process of "creation" involved here, for example, the creation of institutions and statuses, is not some magical power of conjuring up a reality that is independent of the discourse involved. The central point is that there is no independent object of discourse. The discourse and the object of the discourse are one and the same.

The Model

The model is sometimes called the *self-referential model* and sometimes the *performative model*. The latter label goes back to J. L. Austin's account of what he called *performative utterances*—that is, utterances such as "I greet you" or "I curse you" or, when said by an appropriate authority, "I declare you husband and wife." In each case, the act of utterance makes true the thing that is uttered, so its truth does not reside in a correspondence of the utterance with an external reality. It should perhaps be emphasized that in Barnes's case, as in Austin's, this form of analysis in no way *challenges* the commonsense assumption that these self-referential processes all take place within a material setting, that is, a discourse-independent reality whose empirical features can be the object of enquiry, reference, and physical engagement. Both Barnes's and Austin's versions of performative analysis presuppose that it is the environment outside the circle of self-reference that enables the circle to be established in the first place. Barnes calls this the process of *priming* the self-referential system.

Avoiding Misconceptions

The performative or self-referential model captures the obvious truth that social reality is a reality that humans *create* for themselves by virtue of their interacting with one another. But the model does more than confirm a truism: It also *explains* the mysterious character of social ontology. We hear a word that refers to an aspect of social reality, but when we look to see what it refers to, we cannot find anything. The strangeness derives from the false expectation that the word corresponds to some object that is independent of it, when there is no such object apart from other uses of the word. Failure to understand the mechanics of self-referential discourse may tempt the unwary to postulate mythical pseudo-objects to play the role that the empirical object plays in the case of N-predicates. Durkheim argued that the error of assimilating self-referential and other-referential discourse lies at the bottom of all religious and metaphysical speculation. Gods and spirits, as well as the mysterious powers attributed to the human soul and mind, are, in reality, misunderstandings of social processes. They are the transfigured experience of the social group acting upon itself.

Reflexive Awareness: The Mark of the Social Sciences

There is a venerable philosophical tradition in which the human sciences are differentiated from the natural sciences on the grounds that human beings possess special mental characteristics such

as self-awareness. To engage in any social action involves the actors thinking of themselves as being engaged in these actions. Thus, to give a gift involves thinking that you are giving a gift. Where does Barnes's model stand in relation to this tradition? The performative model permits a sharper formulation of these cloudy insights and avoids some of the logical pitfalls to which traditional formulations are subject. Indeed, one of the surprising consequences of Barnes's work is to expose a hidden individualistic premise in the traditional standpoint.

To see this, note that if gift giving really involved thinking that you are giving a gift, it would be impossible to define giving a gift. As Anscombe pointed out, to define "gift giving," it would be necessary to mention thinking that one is giving a gift, but then it would be necessary to define "thinking that one was giving a gift," and that would involve mentioning gift giving—which takes us back to the point from which we started. The definition would run in a circle. Barnes's account avoids this trap. The trap arises because in the traditional picture, self-awareness and self-reference are conflated. The process of collective self-reference, which is central to Barnes's model, has been misconstrued. A social process has been treated as an individual mental process and attributed to the inner structure of a mysterious "intentional state." Barnes takes the process of self-reference out of the head and puts it back where it belongs, that is, in the realm of collective behavior. The only form of self-awareness necessary on Barnes's model is one that enables individual actors to be susceptible to the causal influences exerted by others in the course of interaction. Thus, the only "thinking" needed to make an act an act of gift giving is that of being a responsive participant in the social institution of gift giving.

Critical Points

On first encounter, Barnes's model of social institutions may seem open to the objection that it makes society a mere matter of words. Slaves are slaves because they are called "slaves"—isn't that the theory? Barnes seems to have ignored the harsh, objective reality of exploitation and abuse. Being of low or marginal status can be a matter of life and death. If this objection were right, it would be devastating, but Barnes's model is not just about words, although the use of words has played a central role in its exposition. To be *called* a slave stands duly for all modes of treating someone as a slave and behaving toward them as a slave. Properly understood, the model does *not* minimize the way real-world consequences flow from, and motivate, the processes of social classification and the operation of social institutions. The model is simply designed to capture the essential character of social entities, such as institutions, and to say what they are, regardless of whether their consequences are benign or malign. Indeed, the model can help us understand the processes by which people sustain institutions that systematically deny others access to the resources they need. One can be sure that those who benefit from institutions that embody social injustice will be only too happy to encourage false and mystifying theories about the nature of social reality. The performative model can help dispel the mystification.

David Bloor

See also Institutions as Moral Persons; Kinds: Natural Kinds Versus Human Kinds; Searle and the Construction of Social Reality; Social Construction of Reality; Social Constructivism; Social Ontology, Recent Theories of; Speech Acts

Further Readings

Anscombe, G. E. M. (1981a). On promising and its justice, and whether it need be respected in foro interno. In *The collected papers of G. E. M. Anscombe* (Vol. 3, chap. 2). Oxford, England: Blackwell. (Original work published 1969)

Anscombe, G. E. M. (1981b). Rules, rights and promises. In *The collected papers of G. E. M. Anscombe* (Vol. 3, chap. 2). Oxford, England: Blackwell. (Original work published 1978)

Barnes, B. (1983). Social life as bootstrapped induction. *Sociology, 17,* 524–545.

Barnes, B. (1988). *The nature of power.* Cambridge, England: Polity Press.

Barnes, B. (1995). *The elements of social theory.* London, England: UCL Press.

Personal Identity and Trauma

Many survivors of traumatic events have reported that their selves were shattered by the trauma. Survivors of Nazi death camps have said that they died in the camps, war veterans have remarked

that they lost themselves in battle, and rape victims have written of missing their lost selves. This entry attempts to make sense of these cryptic claims by discussing the effects of trauma on the self and exploring the connections between the philosophical literature on personal identity and the psychological literature on recovering from trauma. To the extent that social science deals with social persons, and hence with notions of selfhood, the study of traumatic experiences that shatters unified conceptions of selfhood is as crucial to social science as it has been to philosophers discussing personal identity over time.

Philosophical Accounts of Personal Identity

The question of personal identity, as addressed by philosophers for centuries, has traditionally been framed as "What are the conditions that make it possible (or impossible) for the same person to persist over time?" The various answers philosophers gave to this question were taken to be revealed by our intuitive responses to thought experiments, such as the one proposed by the 17th-century British philosopher John Locke, in which the consciousness of a prince is transferred from the prince to the body of a cobbler while the consciousness of the cobbler comes to inhabit the body of the prince. "Who is (now) the prince?" Locke asked. Locke, who defended a psychological criterion of personal identity, according to which what makes someone the same person over time is that person's consciousness, answered, "The one with the cobbler's body." Other personal identity theorists who have defended a bodily or biological criterion of personal identity have answered, "The one with the prince's body."

Philosophers since Locke have continued to puzzle over whether persons can survive transformations of various kinds. In recent decades, personal identity theorists have come up with increasingly imaginative and high-tech thought experiments involving freezing and thawing, dissolution and reconstitution, fission, and teletransportation, designed to test our intuitions about what makes someone the same person over time. They have not, until very recently, however, paid attention to the actual experiences of real people who claim to have been utterly transformed by trauma. This may be explained by the fact that what philosophers have traditionally sought in a theory of personal identity is an account of what makes it possible to reidentify a person at one time as the same person that existed at another time. Philosophers have sought a criterion of numerical identity for persons over time, a criterion different from, but with the same function as, a criterion of numerical identity for things, such as trees and ships and bicycles. If, in contrast, we want a theory of personal identity to be an account of a person's self-conception over time, then the study of trauma, including the examination of first-person narratives of trauma survivors, becomes not only relevant but also necessary to the search for a theory with adequate explanatory power.

Psychological Accounts of Trauma

Following the psychiatrist Judith Herman, this entry defines trauma as an experience of being completely helpless in the face of an overwhelming, life-threatening force. Whether it results from human-inflicted violence or a natural disaster, trauma overwhelms a survivor's agency and cognitive capacities and can lead to posttraumatic stress disorder, with symptoms that include hypervigilance, heightened startle response, sensory flashbacks, dissociation, and numbing. The intermingling of mind and body in these symptoms highlights the embodied nature of the self and suggests the inadequacy of both strictly psychological and strictly biological criteria of personal identity.

Psychologists and psychiatrists have observed that trauma survivors suffer the severing of bonds of trust with others and the shattering of fundamental assumptions about their safety in the world. The loss of sustaining attachments to others—family, friends, and community—is experienced as a loss of self by many trauma survivors. This phenomenon, along with the essential role reconnection to others plays in healing from trauma, is taken by some philosophers to support the view that the self exists only in *relation* to other people.

Trauma survivors who lose the ability to identify with their past selves and to project themselves into the future can become paralyzed in a present that no longer makes any sense. In order to carry on, they need to find meaning in the world again; talking about their traumatic experiences in the company of empathetic others has been found to facilitate this. This can be taken to motivate the view that the self is not only relational but also *narrative* in structure, unable to continue without the ability to tell a story about itself. The philosopher Hilde Lindemann

Nelson stresses the importance for survivors of being able to come up with empowering counter-narratives to repair narrative identities damaged by trauma.

As illuminating as the study of trauma can be for understanding the nature of personal identity, it remains unclear just what we ought to make of survivors' claims to have outlived their former selves. Unlike casually made everyday claims such as "I'm not myself today," trauma survivors' claims of lost or destroyed selves appear to be more than mere façons de parler. But if they are literally true, then who is the "I" who is able to say "I am no longer the self I was before the trauma"? What connects the "I" referring to the speaker with the "I" referring to the pre-trauma self? Whether or not these questions can be answered, the conversation between psychologists and philosophers about personal identity and trauma is a lively and engaging one that has only just begun.

Susan J. Brison

See also Collective Identity and Cultural Trauma; Identity, Personal (Philosophy of); Identity, Social; Self and the Social Sciences

Further Readings

Brison, S. J. (2002). *Aftermath: Violence and the remaking of a self*. Princeton, NJ: Princeton University Press.

Herman, J. (1992). *Trauma and recovery.* New York, NY: Basic Books.

Janoff-Bulman, R. (1992). *Shattered assumptions: Towards a new psychology of trauma*. New York, NY: Free Press.

Nelson, H. L. (2001). *Damaged identities, narrative repair.* Ithaca, NY: Cornell University Press.

Pessimistic Induction

A pessimistic induction is an argument that makes use of empirical claims about the history of science—especially accounts of past theories that are false by our lights—to argue that we have little reason to believe that our own theories are even approximately true, or refer to real objects, in the parts of their claims that go beyond matters observable by the human senses. These are claims such as that electrons have spin, since electrons are too small to observe directly with the human eye. The view that we do have reason to believe these things is called *scientific realism*.

What is usually counted as the first argument of this type, given by Larry Laudan, is not strictly an induction. That is, it does not enumerate examples of past failures of science in order to make a positive inference that we have reason to believe that our theories are also false. Rather, since the particular realist views Laudan argued against had put forth strong general claims, such as that the predictive success of our mature theories was a reason to think that they are approximately true, examples of past scientific theories with properties that did not fit those claims could be used to falsify them. Laudan argued that counterexamples to the connection between success and truth are very common in the history of science. We no longer believe the humoral theory of medicine or the physics hypothesis of the existence of an ether-permeating space, but these were assumptions that were successful in prediction.

Objections can be made to the relevance of some examples, though. The humoral theory of medicine enjoyed a reign of many hundreds of years from ancient to medieval times, but its success in diagnosing and curing disease does not seem comparable with that of our germ theory, for example. The only response to the plague that it had available were things like herbs and bloodletting. As for the ether, although physicists believed that there was such a thing, it has been argued that the assumption played no crucial role in their calculations leading to successful predictions. On the other side, it has been argued that the scientists did not know that the ether assumption was unnecessary to the success their physical theory as a whole enjoyed. We similarly may be unaware that some parts of our successful theories are inessential to their success. Thus, even if success were an indicator of truth, we may not know which truths our theories' success indicates.

Pessimistic arguments that have taken the form of a simple induction—roughly, since most successful theories in the past were false, our successful theories now probably are too—have fallen on hard times with the realization that these judgments depend on base rates. For the probability that a theory is true given that it is successful depends not only on the probability that a theory will be successful given that it is true but also on the probability that a theory chosen randomly without regard for its success will

be true and that a theory chosen randomly without regard for its truth will be successful. Making inferences without all of these probabilities is fallacious, and making an accurate judgment of what they are seems impossible.

The potent challenge that the history of scientific failures presents to us is better formulated by John Worrall, who focuses on the observation that for a given currently accepted theory, there is usually a theory that was accepted in the past and that contradicts our theory. The past scientists had evidence just as we do, so what reason can we give for thinking that we will not suffer the same fate of having our theories replaced with future theories that contradict them? A pessimistic argument of a similar form is given by P. K. Stanford and points to the fact that there were alternatives to our predecessors' theories, alternatives that were not conceived by them. We know this because we have since conceived of some of them, and they are written into our accepted theories. What is the reason to think that we are not also subject to conceivable but as yet unconceived alternative theories that will replace ours someday?

One kind of successful response to challenges of this form would be to explain how our justification for believing our theories is different in some principled way from the justification our predecessors had for believing theirs. Sherylin Roush has argued that the principled difference is our continual creation of new methods of investigation and evaluation. In a way that is highly relevant to reliability, we are not simply doing the same thing as our predecessors were, and doing more of it. She argues that this makes the failures of our predecessors to formulate true theories largely irrelevant to judging the expected truth or falsity of our own.

Another strategy, that of Worrall, is to accept that there is no reason to think that our theories will not be replaced but to point out that there is reason to think that significant aspects of our theories will be retained in the replacement. The reason is that significant aspects of successful past (fundamental) theories have been retained in ours. Despite the inconsistency of the older theories with ours, the retention of parts of the older in the newer theories means that we can continue to regard the earlier theories as approximately true. Similarly, we have reason to expect that though our theories will be replaced, we will be able to continue to regard them as approximately true. Worrall argues that in fundamental physics, mathematical structure has been retained over several centuries, a thesis that undergirds a view called *structural realism*.

Sherrilyn Roush

See also Abduction and Inference to the Best Explanation; Bayesianism, Recent Uses of; Induction and Confirmation; Realism and Anti-Realism in the Social Sciences

Further Readings

Laudan, L. (1981). A confutation of convergent realism. *Philosophy of Science, 48,* 19–48.

Roush, S. (2010). Optimism about the pessimistic induction. In P. D. Magnus & J. Busch (Eds.), *New waves in philosophy of science* (pp. 29–58). London, England: Palgrave-Macmillan.

Stanford, P. K. (2006). *Exceeding our grasp: Science, history, and the problem of unconceived alternatives.* Oxford, England: Oxford University Press.

Worrall, J. (2009). *Miracles, pessimism, and scientific realism.* Retrieved from http://www2.lse.ac.uk/philosophy/WhosWho/staffhomepages/Publications/NMAandPIfinal.pdf

Phenomenological Schools of Psychology

Phenomenology is a transnational, transdisciplinary approach to the study of conscious experience. Phenomenological research methods were originally articulated and developed by the philosopher Edmund Husserl in epistemology and ontology at the beginning of the 20th century. This approach has been extended into other areas of philosophy such as ethics and has generated knowledge in the full range of social, scientific, and humanities disciplines, from sociology and economics to theology and literary studies. This broad, dynamic movement continues to be vital and productive today.

Phenomenological psychology has been particularly significant for the philosophy of the social sciences in raising epistemological issues about human understanding.

This entry focuses on the phenomenological approach in psychology and related disciplines by delineating, with historical perspective, its

methodological orientation, fundamental concepts, major proponents, and recent advances.

Phenomenology: History and Basic Tenets

Above all, phenomenological research is based on a fresh and direct contact with the subject matter, apart from received knowledge, as it presents itself in living experience. Following from Edmund Husserl's inspiring and often cited call to return *Zu den Sachen selbst* (to the things, or matters, themselves), phenomenologists employ procedures developed by Husserl (technically called *epochés* [plural of *epoché*], from the Greek ἐποχή, pronounced as "ĕ-pō-ché," meaning "abstention/withdrawal") to set aside prior presuppositions drawn from theories, research findings, hypotheses, and material reality. Because it begins instead with concrete examples of mental life, phenomenology has been characterized by Husserl and later by Herbert Spiegelberg as proceeding "from below" and "from the grassroots." This kind of investigation aims to clarify the essential structure of its subject matter (i.e., to discover its essence, *eidos* in Greek) by using a procedure of meticulous reflections on specific examples, called "eidetic analysis." Crucial to this procedure, as explicated by Husserl, is the free imaginative variation of examples of the phenomenon and the discernment of the invariant core configuration of constituents and experiential processes that qualify all imaginable examples as instances of the phenomenon under study.

One of Husserl's fundamental insights regarding consciousness is its essential quality of "intentionality," the universal act–object correlation. Conscious acts such as perceiving, remembering, thinking, and feeling always "intend" ("are about") objects, such as a food on a plate, a former home, an equation, and a loved one. Moreover, experience apprehends the meanings of objects, including their usefulness and value, within their larger contexts. This most overarching context of living experience is called the *lifeworld* or *existence*, which includes embodied presence, practical engagement, relations with others, surrounding spatiality, temporality, and collective historicity. This concrete, lived world is also the ground of scientific knowledge. Each discipline, by employing its special attitude, thematizes and conceptualizes its own distinctive province of the lifeworld. Psychology, the discipline that focuses on individual persons and finds them to be centers of experience and situated action in the lifeworld, has been of central concern to phenomenological philosophers and psychologists for more than 100 years.

Phenomenological approaches were implicitly present in the psychological research of philosophers such as Wilhelm Dilthey, Franz Brentano, Søren Kierkegaard, and Friedrich Nietzsche. Husserl, who began to make the implications of the phenomenological approach to psychology explicit, was followed by a host of other philosophers who addressed the general philosophical problems of psychology and carried out specific psychological investigations, including Martin Heidegger, Jean-Paul Sartre, Simone de Beauvoir, Maurice Merleau-Ponty, Edith Stein, Max Scheler, Emmanuel Levinas, Gabriel Marcel, Alfred Schütz, Gaston Bachelard, and Paul Ricoeur. Topical areas included perception, behavior, the emotions, imagination, memory, language, personality, social relations, human development, and psychopathology. Phenomenological approaches were also implicitly present in the work of psychologists and psychiatrists such as William James, Carl Stumpf, Sigmund Freud (and other psychoanalysts), Carl Jung, Max Wertheimer (and the Gestalt psychologists), Gordon Allport, Carl Rogers, Abraham Maslow, and Kurt Goldstein. Psychologists who explicitly drew on the phenomenological tradition include Karl Jaspers, Ludwig Binswanger, R. D. Laing, Frantz Fanon, Eugene Minkowski, Erwin Straus, Frederick Buytendijk, Eugene Minkowski, Medard Boss, J. H. van den Berg, and Viktor Frankl, to name only a few.

Phenomenological Psychology

Herbert Spiegelberg has provided the most comprehensive account of the early history of phenomenology's influence on psychology and psychiatry. He showed how phenomenological psychology has brought to light new psychological phenomena, provided original insights and understandings, developed new hypotheses founded on concrete analyses of experience, and offered unique tools for scientific investigation. Although early attempts to institutionalize phenomenological psychology were made in the Netherlands and elsewhere, the most deliberate and systematic development of phenomenological psychology began in the 1960s at Duquesne University in the United States. Drawing on the work of phenomenological philosophers, the "Duquesne circle" produced original phenomenological investigations

and educated a generation of researchers who employed the phenomenological approach across the full spectrum of psychological subject matter, including perception and learning (Amedeo Giorgi), personality and psychopathology (William Fischer), social psychology (Rolf Von Eckartsberg), psychotherapy (Anthony Barton), psychological assessment (Constance Fischer), the imagination (Edward Murray), and group psychology (Frank Buckley). Of most lasting significance has been the development of phenomenological research methods for psychology by Amedeo Giorgi, who has specified procedures of collecting and analyzing lifeworld descriptions of psychological phenomena that have been employed internationally by psychologists with a wide spectrum of subject matter. The *Journal of Phenomenological Psychology*, founded by Giorgi in 1970, has been the leading venue for the phenomenological movement and in introducing qualitative research methods in psychology.

Amedeo Giorgi developed a research method for psychology that employs the traditional tenets of phenomenology drawn from Husserl, such as the *epoché* of the natural sciences, a partial *epoché* of the natural world, the focus on the meanings and experiential processes through which the situation presents itself, and the eidetic analysis of psychological structures using imaginative variation. This method distinctively enables researchers to investigate the full range of other people's psychological life as it occurs in the lifeworld. Data are constituted by descriptions of specific situations through which the research phenomenon was lived. Research participants describe their own or other people's experiences, either in written form or in an interview. For instance, the researcher may ask participants to describe a situation in which they learned something, were jealous of someone, or encountered someone who showed "courage."

Giorgi specified procedures that ensure rigorous analysis of all data in the achievement of general psychological conceptualizations of subject matter. First, the description is read over openly. Second, expressed changes in meaning are used to differentiate analyzable "meaning units" in the data. Third, each meaning unit is reflected on and imaginatively varied in order to articulate what it reveals about the matter under investigation and its relevance to the research question. Finally, reflective analyses of numerous descriptions, supplemented by free imaginative variation, are employed in explicating the essential psychological structure, or typical structures, of the phenomenon under investigation.

One important procedure demanded by the last step is the determination of whether the data reveal a single general structure or numerous typical structures. In *Phenomenology and Psychological Research*, using five examples/descriptions to demonstrate this method with the phenomenon of "learning," Giorgi (1985) found two different structures of learning, one never before reported in the literature, and thereby demonstrated how phenomenological research can reveal overlooked psychological processes in an important mainstream area of research. One typical structure of learning had already been identified by traditional psychologists and called "skill acquisition." This kind of learning involves no cognitive ambiguity and changing assumptions but rather a movement from awkward bodily activity to smooth functioning within a single expectational set. However, a second typical structure of learning that Giorgi called "Discovery of Discrepancy Between Assumptions and Situation" was found in a person's encounter with hard facts that are discordant with his or her assumptions. Learning of this kind occurs when the person pauses and thinks about these discordant facts according to a new perspective that illuminates the situation in unexpected ways. A newly clarified ambiguity transforms not only the meanings of the current situation but also, by implication, those of future situations of the same type. Giorgi recognizes both as types of learning in that the person makes "a discovery about himself or his way of relating to the world in such a way that all future situations of that type will be handled in a more adequate way" (p. 66).

In the 2009 demonstration investigating "jealousy" in *The Descriptive Phenomenological Method in Psychology*, Amedeo Giorgi and Barbro Giorgi independently and reliably discerned a single structure in the collected examples and their imaginative variations. They delved into the personal experience of not receiving sufficient attention and appreciation from a desired other person, whose appreciation is actively robbed by a third person. They elaborate the intense feelings of resentment and hostility when this third takes advantage of her or his unfairly privileged position in order to undermine the jealous person's position and possibility of gaining the appreciative attention sought. They also bring out the significance of shame and how, within this general structure,

the jealous person hides her or his experience from others. This demonstration shows the reliable way this kind of analysis leads to convergent results, in this case in the single-structure solution, on the parts of different researchers.

Phenomenology has recently been utilized and the methods articulated by a number of other psychological researchers. Max Van Manen in Canada, Clark Moustakas in the United States, and Jonathan Smith in Great Britain have also developed variant phenomenological research methods for psychology that have been and continue to be used in training and guiding researchers. Consistent with the principles of science, phenomenological psychology offers an evidentiary basis to criticize prior knowledge and to establish new rational claims that are themselves corrigible through methodical extensions of its procedures.

Phenomenological psychologists have generated original knowledge in and clarified relations among nearly every subdiscipline of psychology, including cognitive, social, developmental, experimental, psychometric, personality, behavioral, psychoanalytic, and humanistic. The phenomenological approach continues to be central in psychology's qualitative methodological revolution in the early 21st century.

Noteworthy among the current, continuing contributions to clinical psychology is the phenomenological participatory action research on recovery in schizophrenia by Larry Davidson. With individuals long considered by traditional investigators to be beyond the scope of empathic understanding, Davidson has consistently demonstrated that the lived world of the person with schizophrenia is not only intelligible but highlights the person's own active role in the recovery process. Davidson's phenomenological psychology has articulated key dimensions of this psychopathology beyond the usual focus on deficits, dysfunctions, and etiological theories, including individuals' hopes, assets, agency, and efforts at building a meaningful life in the community. Specifically, he details not only their "symptoms" judged "abnormal" by others but also their suffering of rejection from loved ones and friends, as well as their struggles to establish gainful employment, meaningful family involvements, confidence in navigating the world, contributions to others and society, and a hopeful future. Guided by transcendental phenomenology, Davidson's research on recovery in psychopathology integrates interdisciplinary work in demonstrating that what psychosis is cannot be understood apart from the lived world of stigma, poverty, unemployment, political oppression, and social marginalization. These analyses, based on in-depth interviews, have led to the development of community programs that employ the resources of peers and natural supports to empower recovering persons as social agents in their communities.

Concluding Remarks

Phenomenological psychology has entered the 21st century as a continuing movement, self-aware and self-critical. Phenomenological psychologists eschew dogmatism and exclusivity as they emphasize evidence and reflection, diversity and ambiguity, vision and revision. They aim to be continual beginners, gaining new knowledge by setting aside prior knowledge and returning to freshly reflect on concrete examples of living experience.

Frederick J. Wertz and Miraj U. Desai

See also Being-in-the-World; Embodied Cognition; Existential Phenomenology and the Social Sciences; Existential Psychology; Gestalt Psychology; Hermeneutics, Phenomenology, and Meaning; Intentionality; Life-World; Schizophrenia: Psychoanalytic, Phenomenological, and Contemporary Philosophical Approaches

Further Readings

Davidson, L. (2003). *Living outside mental illness: Qualitative studies of recovery in schizophrenia*. New York, NY: New York University Press.

Davidson, L., Staeheli, M., Stayner, D., & Sells, D. (2004). Language, suffering, and the question of immanence: Toward a respectful phenomenological psychopathology. *Journal of Phenomenological Psychology, 35*(3), 197–232.

Giorgi, A. (1970). *Psychology as a human science: A phenomenologically based approach*. New York, NY: Harper & Row.

Giorgi, A. (Ed.). (1985). *Phenomenology and psychological research*. Pittsburgh, PA: Duquesne University Press.

Giorgi, A. (2009). *The descriptive phenomenological method in psychology: A modified Husserlian approach*. Pittsburgh, PA: Duquesne University Press.

Giorgi, A., Barton, A., & Maes, C. (Eds.). (1983). *Duquesne studies in phenomenological psychology* (Vol. 4). Pittsburgh, PA: Duquesne University Press.

Halling, S., & Nill, J. D. (1995). A brief history of existential-phenomenological psychiatry and psychotherapy. *Journal of Phenomenological Psychology, 26*(1), 1–45.

Husserl, E. (1982). *Ideas: General introduction to pure phenomenology* (W. R. B. Gibson, Trans.). New York, NY: Collier. (Original work published 1913)

Husserl, E. (1977). *Phenomenological psychology: Lectures, summer semester, 1925* (J. Scanlon, Trans.). Boston, MA: Martinus Nijhoff. (Original work published 1925)

Spiegelberg, H. (1972). *Phenomenology in psychology and psychiatry.* Evanston, IL: Northwestern University Press.

Valle, R. (1998). *Phenomenological inquiry in psychology: Existential and transpersonal dimensions.* New York, NY: Plenum Press.

Wertz, F. J. (2006). Phenomenological currents in twentieth-century psychology. In H. L. Dreyfus & M. A. Wrathall (Eds.), *A companion to phenomenology and existentialism* (pp. 394–411). Oxford, England: Blackwell.

Wertz, F. J., Charmaz, K., McMullen, L. M., Josselson, R., Anderson, R., & McSpadden, E. (2011). *Five ways of doing qualitative analysis: Phenomenological psychology, grounded theory, discourse analysis, narrative research, and intuitive inquiry.* New York, NY: Guilford Press.

Philosophes, The

Although the French term *les philosophes* means philosophers in the strict sense, it also has a broader meaning for which there is no exact equivalent in English, hence its widespread use in other languages to refer to the group of public intellectuals and writers in 18th-century Europe and the United States associated with the movement now known as the Enlightenment, very few of whom were philosophers in the more restricted sense in which that word tends to be used in English today. They included natural scientists, philosophers, clerics, politicians, public officials, novelists, and journalists with a shared commitment to reason, science, and religious toleration.

While the best-known philosophes were French—men such as Voltaire, Denis Diderot, Jean le Rond d'Alembert, the Comte de Buffon, Étienne Bonnot de Condillac, the Baron de Montesquieu, and Jean-Jacques Rousseau, who was a French-speaking citizen of Geneva—they had counterparts across Europe and the United States, most notably Adam Smith and David Hume in Scotland, Immanuel Kant and Gotthold Ephraim Lessing in Germany, Thomas Jefferson and Benjamin Franklin in the United States, and Cesare Beccaria in Italy. These thinkers commonly referred to themselves as a society of men of letters ('*société des gens de lettres*') who owed allegiance to an international "Republic of Letters" devoted to the popular dissemination of enlightenment and the promotion of reform that began around the middle of the 18th century and continued until the French Revolution in 1789.

Most of the philosophes were empiricists who rejected the belief in original sin and the concept of innate ideas, both of which were closely associated with orthodox Christianity. They typically viewed the human mind as a blank slate that acquired knowledge via sensory experience. The *locus classicus* of this view is John Locke's 1690 *Essay Concerning Human Understanding*, which was widely read and admired throughout the 18th century by the philosophes. They attacked mysticism and obscurity in favor of clarity and openness, and they were champions of the scientific method of experimentation and direct empirical observation of nature pioneered in the 17th century by men such as Francis Bacon. Most of the philosophes were confident that by this approach we could understand the natural and human worlds, which was the surest means to progress and happiness.

The epitome of the philosophes' ambition to spread enlightenment as broadly as possible was the epic and ambitious *Encyclopédie*, edited by Diderot and d'Alembert. Published between 1751 and 1765 in 17 large volumes, it consisted of tens of thousands of entries and illustrations (with several later supplements and editions) written by virtually all of the leading philosophes of the day in France on a huge range of topics, from music and natural science to politics and religion. This controversial project was officially banned in France in 1759, although it enjoyed the support of some very high-placed officials, including the director of censorship and publication, Guillaume-Chrétien de Lamoignon de Malesherbes, who worked surreptitiously to ensure that the *Encyclopédie* continued to be published unofficially.

Among the most common myths about the philosophes is that they were naive optimists who believed in the inevitability of progress, even though Voltaire, the quintessential Enlightenment figure, openly mocked this view in his popular novel *Candide* (1759). At the other extreme was the Marquis de Condorcet's wildly optimistic *Sketch for a Historical Picture of the Progress of the Human Mind*, published in 1795 shortly after his death in prison during the French Revolution. The mainstream view of the philosophes on progress lay somewhere between these two extremes, tending toward a very cautious optimism about the prospects for improvement, with a keen sense of how slow and uncertain it could be. Even so, most believed that things had gradually improved and would likely continue to do so as reason, toleration, and science displaced religious fanaticism, intolerance, and superstition.

Another common misconception of the philosophes is that they regarded society as unnatural. In fact, most of them criticized social contract theory and affirmed natural human sociability, which they practiced in salons, cafés, and academies, where they met to debate and exchange ideas.

Although often caricatured by their opponents as atheists, very few philosophes denied the existence of God. Most were deists who rejected Christianity, attacked religious persecution, mocked traditional religious institutions and beliefs, and endorsed a minimalistic natural religion. With few exceptions, they were anticlerical, were highly critical of religious fanaticism, and advocated religious toleration. While some philosophes, such as Claude Adrien Helvétius and the Baron d'Holbach, rejected belief in God, they were rare exceptions, and many were as critical of atheism as they were of religious fanaticism.

In most European countries in the 18th century, the *philosophes* saw themselves as a self-appointed, unofficial opposition to the dominant political establishment. In some rare cases, they actually held political power, as in the United States (where Thomas Jefferson became president) and Prussia under Frederick II. There was surprisingly little agreement among them on the best means to the ends of social reform and the promotion of human well-being, which all agreed should be the purpose of government. Some, like Voltaire and d'Alembert, favored enlightened despotism, while others (a minority), like Rousseau, were democrats who put their faith in the virtue and good sense of ordinary people.

Graeme Garrard

See also Empiricism; Encyclopedia; Enlightenment, Critique of; Montesquieu and the Rise of Social Science; Social Contract Theories

Further Readings

Cranston, M. (1986). *Philosophers and pamphleteers: Political theorists of the Enlightenment*. Oxford, England: Oxford University Press.

d'Alembert, J. (1963). *Preliminary discourse to the encyclopedia of Diderot* (R. Schwab, Trans.). Indianapolis, IN: Bobbs-Merrill.

Gay, P. (1966–1969). *The enlightenment: An interpretation* (2 vols.). New York, NY: Vintage Books.

Hyland, P., Gomez, O., & Greensides, F. (Eds.). (2003). *The Enlightenment: A sourcebook and reader*. New York, NY: Routledge.

Israel, J. (2001). *Radical enlightenment: Philosophy and the making of modernity, 1650–1750*. New York, NY: Oxford University Press.

Kant, I. (1991). What is enlightenment? In H. S. Rei (Ed.), *Kant: Political writings* (H. B. Nisbet, Trans.). Cambridge, England: Cambridge University Press.

Kors, A. (1976). *D'Holbach's coterie: An enlightenment in Paris*. Princeton, NJ: Princeton University Press.

Outram, D. (2006). *The Enlightenment*. Cambridge, England: Cambridge University Press.

Schmidt, J. (Ed.). (1996). *What is enlightenment? Eighteenth-century answers and twentieth-century questions*. Berkeley: University of California Press.

Voltaire. (1972). *Philosophical dictionary* (T. Besterman, Trans.). London, England: Penguin.

Philosophical Psychology, History of

Philosophical psychology is the philosophical analysis of theoretical psychological constructs such as rationality, emotion, motivation, consciousness, and intelligence. This entry briefly describes its first recognition as a discipline distinct from scientific or empirical psychology and its brief flourishing as a distinctive philosophical program in the 20th century.

As Hermann Ebbinghaus said of psychology, philosophical psychology has a long history but a short past. While philosophers have critically reflected on the theoretical constructs employed to explain human psychology and behavior for millennia, philosophical psychology was only recognized as a discipline distinct from scientific or empirical psychology in the 18th century, and it only came to fruition as a distinctive philosophical program for a relatively short period in the latter half of the early 18th century.

Since at least the time of the ancient Greeks, philosophers have engaged in critical conceptual analysis of the theoretical constructs employed in the explanation of human psychology and behavior, while advancing tentative theories of human psychology and behavior based (however tentatively) upon observation (including introspective observation) and experiment. Thus, for example, in his discussion of anger in *De Anima*, Aristotle not only describes the physiological basis of anger as the boiling of the blood around the heart (on doubtful empirical grounds) but also explicates the conceptual content of anger as an appetite for returning pain for pain.

Although many ancient, medieval, Renaissance, and Enlightenment thinkers provided quite sophisticated conceptual analyses of theoretical psychological constructs, these were always mixed with theoretical speculation and appeal to empirical facts; and the notion of a form of philosophical psychology distinct from empirical psychology did not develop until the 18th century, when the German philosopher Christian Wolff distinguished between rational and empirical psychology. According to Wolff, rational psychology is concerned with the rationally demonstrable principles of psychology (relating to the human soul), while empirical psychology is concerned with the empirical description and measurement of psychological faculties such as sensation, memory, and intellect.

Immanuel Kant demonstrated that there are no rationally demonstrable principles of psychology analogous to those of logic and mathematics (attacking the hyperbolic claims of reason), on the one hand, and that empirical psychology cannot attain the status of a genuine quantified science, on the other—a claim supposedly refuted by the development of psychophysics in the late 19th century (although Kant did not deny the quantitative measurement of the intensity of sensation but only the possibility of a quantified dynamical psychology analogous to Newtonian physics).

Nevertheless, most post-Kantian theorists ignored the distinction and continued to advance mixed philosophical, theoretical, and empirical claims about human psychology and behavior even after analytic philosophers and phenomenologists distinguished themselves from empirical psychologists in the early 20th century as a consequence of their rejection of psychologism, the view that the laws of thought are a part of empirical psychology. While they distinguished themselves from empirical psychologists, both analytic philosophers and phenomenologists conceived of psychology as a scientific inquiry, concerned with the causal explanation of human psychology and behavior.

This conception of psychology was challenged by a number of philosophical psychologists in the 1950s and 1960s, who denied that explanations of meaningful human action in terms of reasons or motives could be treated as analogous to causal explanations in the natural sciences. These philosophers, many of whom were students or followers of Ludwig Wittgenstein, maintained that reason or motive explanations violate David Hume's dictum that causes and effects are contingently related, because there is a "conceptual connection" (or internal relation) between reasons or motives and the actions they explain—for example, between motives of revenge and acts of revenge. Their distinctive position on this question reprised earlier 19th- and 20th-century views in the philosophy of social science, which maintained that there is a fundamental distinction between the modes of (causal) explanation and understanding characteristic of the natural and the human sciences. More generally, these philosophers endorsed Wittgenstein's famous claim that "in psychology there are experimental methods and conceptual confusion" (*Philosophical Investigations,* Pt. II, p. xiv) and developed detailed conceptual analyses of psychological constructs such as perception, emotion, motivation, intention, the unconscious, voluntary action, rationality, and self-deception. Seminal works in this tradition included G. E. M. Anscombe's *Intention* (1958), R. S. Peters's

The Concept of Motivation (1958), Anthony Kenny's *Action, Emotion and Will* (1963), Alasdair MacIntyre's *The Unconscious: A Conceptual Analysis* (1958), A. I. Melden's *Free Action* (1961), Norman Malcolm's *Dreaming* (1959), and Peter Winch's *The Idea of a Social Science* (1958). With the exception of Anscombe's *Intention*, all these works were published in the Routledge and Kegan Paul series *Studies in Philosophical Psychology*, later known as the "little red books" because of their distinctive color and size. This, predominantly British movement was relatively short-lived, and its death knell was sounded by Donald Davidson's 1963 article "Actions, Reasons and Causes," which maintained that reason explanations are a species of ordinary causal explanations and repudiated the "conceptual connection" argument advanced by the "Red Book Philosophers."

Although most philosophers have abandoned arguments against the causal explanation of meaningful human action in terms of reasons and motives, they have continued to develop conceptual analysis of the theoretical constructs of psychology, even as the status of conceptual intuitions has been questioned as philosophy itself has become more empirical and experimental and conceptual analysis has come to be treated as continuous with theoretical and empirical psychology. In recent years, the conceptual analysis of theoretical psychological constructs has been subsumed within the philosophy of psychology, which is nowadays frequently equated with philosophical psychology, as, for example, in the journal *Philosophical Psychology*, one of the premier journals in contemporary philosophy of psychology.

John D. Greenwood

See also Action, Philosophical Theory of; Causes Versus Reasons in Action Explanation; Experimental Philosophy; Explanation Versus Understanding; Mind–Body Relation; *Naturwissenschaften* Versus *Geisteswissenschaften*

Further Readings

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy, 23,* 685–700.

Melden, A. I. (1961). *Free action.* London, England: Routledge & Kegan Paul.

Peters, R. S. (1950). *The concept of motivation.* London, England: Routledge & Kegan Paul.

Richards, R. (1980). Christian Wolff's prolegomena to empirical and rational psychology: Translation and commentary. *Proceedings of the American Philosophical Society, 124*(3), 227–239.

Wittgenstein, L. (1973). *Philosophical investigations* (3rd ed.; G. E. M. Anscombe, Trans.). New York, NY: Pearson.

Philosophy of Economics, History of

The philosophy of economics, broadly understood as the philosophical reflection on economic matters, is almost as old as Western philosophy itself and dates back to the works of the ancient Greek thinkers Xenophon and Aristotle. More useful is a narrower understanding of the term as the systematic investigation of the nature and methods of the science of economics, which is contingent on the birth of economics (or political economy, as it was known then) as a discipline in the 18th century. The English philosopher and polymath John Stuart Mill was arguably the first philosopher of economics in this narrower sense. This entry surveys the development of ideas concerning the nature and methods of economics from its Millian origins until the present day.

The Methodological Tradition: John Stuart Mill and Neville Keynes

John Stuart Mill (1806–1873) is widely known for his writings on logic and the philosophy of science, utilitarianism and liberty, feminism, and classical economic thought. As a philosopher of economics, his major contributions are three interrelated ideas: the characterization (1) of economics as an *abstract science*, (2) of its method as *a priori*, and (3) of causal laws as *tendencies*.

Mill (1830/1948) defined economics as

> the science which traces the laws of such of the phenomena of society as arise from the combined operations of mankind for the production of wealth, in so far as those phenomena are not modified by the pursuit of any other object. (p. 140)

The most notable features of this definition are (a) that economics is defined in terms of the causes responsible for certain phenomena and (b) that these phenomena are conceived of as hypothetical: whatever *would* follow from the human pursuit of wealth if no other motive *were* present. This is what Mill meant by economics being an "abstract science": For the most part it does not describe concrete states of affairs we can see, touch, and feel because these are normally the result of a concomitance of causes. Instead, it describes what happens "in the abstract" when noneconomic factors are absent.

Concerning method, Mill distinguishes two main approaches: the *a posteriori* and the *a priori* methods. The former proceeds from specific experiences to a general conclusion, and thus inductively, while the latter, by reasoning from an assumed highly general hypothesis (which is supported by a wide range of experiences) to a specific conclusion, therefore mixing inductive and deductive argument. Mill thought that the nature of social phenomena undermines attempts to employ the a posteriori method fruitfully. Economists cannot perform experiments, nor do economic phenomena resemble natural situations that can be sufficiently captured by experiments, since such phenomena are too varied. Therefore, the economist must resort to the a priori method. She is, however, helped by the fact that the desires of human beings and what triggers them are observable and thus that the relevant laws are known. The main difficulty for the economist is therefore to calculate what will happen when the laws of human nature operate in a specific situation, but this is, according to Mill, not part of the business of science but of its application.

Causal laws, whether concerning humans or natural phenomena, are to be understood, according to Mill, as tendencies. That is, causal laws describe not what actually happens but rather what tends to happen in the absence of disturbing causes. In economics, these disturbances are the noneconomic causes—everything apart from the pursuit of wealth. Mill thought that in economics, different causes combine "mechanically" rather than "chemically." That is, when the operation of one cause is disturbed by another, both causes continue to affect the result in the same direction as they would operate if the other cause were absent. In chemistry, by contrast, when two substances combine, their product has completely different properties from those of its components. The success of the a priori method depends on the truth of this principle of composition: Understanding what economic factors do "in the abstract" would be of no use unless they contributed to results in predictable ways when other factors are present.

John Neville Keynes (1852–1949) was an economist and the father of John Maynard Keynes. In his methodological writings, he tried to adjudicate in the *Methodenstreit*, between the German historical school and the Austrian school of economics, by combining inductive and deductive elements more rigorously than Mill had. His attempts were met with little success, however, as subsequent generations of economists, especially after World War II, paid little attention to evidence in the choice and formulation of premises from which conclusions about concrete economic phenomena were to be derived. A lasting contribution to methodology is Keynes's three-part distinction of *positive*, *normative*, and *applied* economics. His notion of positive economics as a "body of systematized knowledge concerning what is" and of normative economics as "a body of systematized knowledge relating to criteria of what ought to be, and concerned therefore with the ideal as distinguished from the actual" (Keynes, 1890/1999, *The Scope and Method of Political Economy*, p. 22) are still in use today. By contrast, *applied economics* to Keynes was synonymous with the "art of economics" and meant more than just the use of economic principles in concrete contexts. It was rather a branch of economics, separate from both positive and normative inquiry, dealing with rules for attaining given practical purposes. It is only in recent years that methodologists have started to pay attention again to this aspect of economic thought.

A Century of Isms: Positivism, Instrumentalism, Falsificationism, and Postmodernism

The main methodological ideas that surfaced during the greater part of the 20th century were influenced by either logical positivism or critical rationalism, or both, or they developed as a reaction to these schools of thought. One uniting feature was their understanding of economics as a body of theoretical

statements such as *assumptions*, *hypotheses*, *postulates*, or *conjectures*.

Two methodological treatises received much attention during the 1930s: Lionel Robbins's *Essay* of 1935 (a first edition had appeared in 1932) and Terence Hutchison's work of 1938. Robbins defined economics as a science that studies behavior as a relationship between ends and scarce means, and he rejected, with Mill and Keynes, the idea that economic principles could be based on specific experiences or controlled experiments. But unlike Mill and Keynes, who aimed to develop sophisticated combinations of inductive and deductive styles of reasoning, Robbins (1935) downplayed the inductive part of the endeavor as the basic postulates of economics "are so much the stuff of our everyday experience that they have only to be stated to be recognized as obvious" (p. 79). Hutchison's book can be understood as a detailed critique of Robbins's ideas from a positivistic standpoint. In particular, he sought to secure economics' status as a genuine science by showing that it is based on substantive principles capable of empirical test rather than mere tautologies, and he emphasized the need for confrontation of these principles with evidence.

Milton Friedman's 1953 essay was and continues to be the most widely discussed contribution to economic methodology. The essay should be read as providing an answer to the question "What kind of evidence is relevant to assessing economic hypotheses?" In the 1940s, doubts were raised about the empirical adequacy of some of the basic postulates of economics on the basis of survey results about how firms make production and pricing decisions. Friedman denied that such evidence should worry economists because economic theories or hypotheses ought to be evaluated on the basis of their predictive success and fruitfulness, not their literal truth or falsehood. This position has later come to be known as instrumentalism, but it also clearly bears positivist, pragmatist, and Popperian influences.

The preponderance of Popperian ideas grew after World War II, and helped define the methodology and philosophy of economics as an independent field of inquiry, especially with the appearance of Mark Blaug's book in 1980. The book emphasized the importance of falsificationism à la Karl Popper and Imre Lakatos and argued that economists often preach falsification, though they rarely practice it.

By the time the book came out, falsificationism (and also positivism) had already largely been given up in philosophy circles because of complications that proved quite recalcitrant, such as the Duhem problem and the difficulty of drawing a meaningful distinction between observational and theoretical statements. In a provocative article and a later book, Donald McCloskey built on these philosophical advances, rejected the 20th-century methodologies as "modernist," and developed an anti-methodological stance named the *rhetoric of economics* to replace them. Instead of using maxims such as "test your hypotheses," "build predictive models," and so on, to evaluate the content of economics, we were invited to attend to the arguments given in support of a position and to assess them using the tools of rhetorical analysis.

The Millian Tradition Revived: Causality, Models, and Evidence

The last decade of the 20th century was marked by two developments in the philosophy of economics. On the one hand, there was a revival of Millian themes, most prominently perhaps in Daniel Hausman's 1992 book, which in many ways built on and developed Mill's methodological ideas. On the other hand, philosophers of economics followed a trend initiated by a group of general philosophers of science to turn away from issues surrounding scientific theory and its assessment and instead to attend more closely to scientific practice. Thus, more applied work in economics moved into the center of methodological attention: econometrics, modeling, experiments, and measurement. In much of this recent work, philosophical analyses of economic practices often also take their cue from Mill. Thus, Nancy Cartwright develops an account of "causal capacities," which is modeled on Mill's notion of a tendency, and she uses it to analyze econometrics as well as models; Uskali Mäki similarly understands economic models as "isolations," which is also closely related to Mill's ideas about abstract science and tendencies; Francesco Guala analyzes economic experiments on the basis of Mill's "canon of inductive methods," and Julian Reiss takes Mill's skepticism about the applicability of inductive methods to economic problems as a foil to develop a more thoroughly evidence-based methodology of economics. We can expect this trend of focusing on more applied matters to continue in the future and also to extend to normative issues such as economic well-being and policy (see,

e.g., the 2009 anthology edited by Harold Kincaid and Don Ross).

Julian Reiss

See also Causation in the Social Sciences; Critical Rationalism; Critical Realism in Economics; Evidence-Based Policy; Experiments in Social Science; Falsifiability; Feminist Economics; Heterodox Economics; Marxist Economics; *Methodenstreit*; Mill and the Moral Sciences; Popper's Philosophy of Science; Realism and Anti-Realism in the Social Sciences

Further Readings

Blaug, M. (1980). *The methodology of economics: Or, how economists explain.* Cambridge, England: Cambridge University Press.

Cartwright, N. (1989). *Nature's capacities and their measurement.* Oxford, England: Clarendon.

Cartwright, N. (1995). Probability and experiments. *Journal of Econometrics, 67,* 47–59.

Cartwright, N. (1999). The limits of exact science, from economics to physics. *Perspectives on Science, 7*(3), 318–336.

Friedman, M. (1953). *The methodology of positive economics. Essays in positive economics.* Chicago, IL: University of Chicago Press.

Guala, F. (2005). *The methodology of experimental economics.* Cambridge, England: Cambridge University Press.

Hausman, D. (1992). *The inexact and separate science of economics.* Cambridge, England: Cambridge University Press.

Hausman, D. (Ed.). (1994). *The philosophy of economics: An anthology* (2nd ed.). Cambridge, England: Cambridge University Press.

Hutchison, T. (1938). *The significance and basic postulates of economic theory.* London, England: Macmillan.

Keynes, J. N. (1999). *The scope and method of political economy.* Kitchener, Ontario, Canada: Batoche Books. (Original work published 1890)

Kincaid, H., & Ross, D. (Eds.). (2009). *The Oxford handbook of philosophy of economics.* Oxford, England: Oxford University Press.

Mäki, U. (1992). On the method of isolation in economics. In C. Dilworth (Ed.), *Idealization IV: Structuralism, intelligibility in science* (pp. 317–351). Amsterdam, Netherlands: Rodopi.

McCloskey, D. (1983). The rhetoric of economics. *Journal of Economic Literature, 21*(2), 481–517.

McCloskey, D. (1985). *The rhetoric of economics.* Madison: University of Wisconsin Press.

Mill, J. S. (1874). *A system of logic.* New York, NY: Harper. (Original work published 1843)

Mill, J. S. (1948). *Essays on some unsettled questions of political economy.* London, England: London School of Economics and Political Science. (Original work published 1830)

Reiss, J. (2008). *Error in economics: Towards a more evidence-based methodology.* London, England: Routledge.

Robbins, L. (1935). *An essay on the nature and significance of economic science.* London, England: Macmillan.

Philosophy of Expertise

Experts are everywhere. People rely on doctors, lawyers, and accountants for many matters of personal well-being. Politicians and managers regularly turn to scientists, professors, and professional consultants for policy advice. Judges and journalists put experts on the spot to weigh in on numerous issues. It is impossible to imagine life without experts. Yet what is an expert?

One answer is that we need to understand what distinguishes expert knowledge from other types of knowledge. Expert knowledge is often understood as *tacit* knowledge of domains like chemistry, the U.S. tax code, or football. Tacit knowledge refers to fluency gained only through long-term immersion in a domain's practices and credentialing processes. That is, tax professionals should not have to constantly reference guides and manuals to do someone's taxes. They should know just how to do it implicitly; the right things should stand out for them. Tax professionals' tacit knowledge can be contrasted with mere memorization of the tax code, which does not produce the fluency needed to make expert judgments.

Though we may accept that expert knowledge is tacit knowledge, we may disagree on who has tacit knowledge suitable for advising others. For example, is an expert in football only someone who has played football at all relevant levels and continued on to a coaching career? Or is a sports journalist who has covered football for many years, but never played competitively, also an expert? The former has had long-term immersion in the practices and credentialing processes of the domain; the latter only has mediated experiences, such as linguistic immersion with players and coaches as well as ample time spent as a spectator. Should both be considered experts?

This question is hard to disentangle from questions about expertise and power relations. Expert status carries authority; authoritative advice influences people's lives. Consider regulatory decisions about risky technologies. Citizens sometimes focus on aspects of risk that experts largely ignore, such as qualitative features that may be independent of probabilistic assessments of harm; they may also have different beliefs about the meaning of safety. In cost–benefit analysis, experts may offer good advice on the costs of permitting and regulating technologies, but they are less helpful in identifying and measuring the benefits. For example, experts' studies of citizens' willingness to pay are often unclear because it is not given that citizens express their values that way or that they know enough about the technologies to determine what they would pay. Should experts even measure people's preferences? Or should they seek to help people to have more informed preferences?

Expert advice can be bent against consumers' interests; expert opinions and decision processes can be used as excuses to sidestep informed consent procedures, transparency, and public or stakeholder engagement. The credentials used to identify experts may be biased in numerous ways. Consider the case of indigenous peoples whose members have traditional ecological knowledge of their lands but no formal scientific credentials. Failure to confer expert status on traditional ecological knowledge disenfranchises Indigenous people from environmental decision making.

The philosophy of expertise also addresses the role of experts in problems such as water availability, sustainable agriculture, and climate change, which are so complex that no one domain of expertise can support the efforts to address them. Climate change, for example, is relevant to domains ranging from meteorology to forestry to animal science to civil engineering to artificial intelligence (for modeling), and many others. Yet insofar as climate change concerns human beliefs, behaviors, norms, visions, and capacities to act, it is also the subject of sociology, cognitive science, literature, anthropology, philosophy, psychology, and art.

There are philosophical issues involved in understanding how experts in different domains ought to work to collaborate on problems like climate change. Experts use different methods (e.g., experimental, fieldwork, philosophical analysis) that are initially incompatible with one another because they reflect different assumptions about the nature of inquiry. Data sets may be difficult to integrate, compare, or even interpret faithfully in relation to one another. Different domains may have ingrained languages, norms, and professional standards that have not been adopted or are looked at critically by experts in other domains. It is also unclear how interdisciplinary collaboration should be assessed. Interdisciplinary collaboration might make problems like climate change even harder to deal with and may be more susceptible to political manipulation.

Efforts to address climate change will also only be effective if research is actionable by citizens in the localities where the findings must be applied. Making research actionable and relevant requires experts to consult nonexperts early on in the planning phases and that nonexperts serve among the external peers who judge the quality of the research. But exactly how far into expert domains should nonexperts go? Should nonexperts shape what research gets funded? Should they be involved in deciding standards of evaluation of evidence and modeling assumptions? Will citizen peer review reduce the autonomy of experts in their own domains?

Interdisciplinarity and nonexpert involvement in research challenge the traditional boundaries of expertise and expert domains such as disciplines and fields. Some philosophers address these challenges directly. They have supplemented typical philosophical work with empirical activities, which include putting together case studies, running experiments, and collaborating with engineers and scientists to set appropriate technical standards for innovations. They work directly with these experts in interdisciplinary research.

Philosophers often bring their expertise into interdisciplinary collaboration. They engage with practitioners of different fields, from biotechnology to organic farming, or with members of different communities, from urban farmers to state policymakers. Philosophers become members of these communities, with their own unique perspectives and contributions. They consider themselves *as* members with special contributions but not as privileged experts. They use this knowledge and status to publish in fora with a wider audience, engaging stakeholders and even the broad public on issues of research ethics and technological risk. The philosophy of expertise properly includes this work because of what it reveals about the value of philosophical expertise in relation to experts in other domains and the public.

Evan Selinger and Kyle Whyte

See also Interdisciplinarity; Evidence-Based Policy; Experimenting Society, The; Policy Applications of the Social Sciences; Risk; Tacit Knowledge; Trust, Epistemic

Further Readings

Collins, H., & Evans, R. (2009). *Rethinking expertise*. Chicago, IL: University of Chicago Press.

Douglas, H. (2005). Inserting the public into science. In P. Weingart & S. Maasen (Eds.), *Democritization of expertise? Exploring novel forms of scientific advice in political decision-making* (Vol. 24, pp. 153–169). Dordrecht, Netherlands: Springer.

Frodeman, R., Klein, J. T., & Mitcham, C. (2010). *The Oxford handbook of interdisciplinarity*. New York, NY: Oxford University Press.

Maclean, D. (2009). Ethics, reasons and risk analysis. In L. Asveld & S. Roeser (Eds.), *The ethics of technological risk* (pp. 115–127). London, England: Earthscan.

McGarity, T., & Wagner, W. (2008). *Bending science: How special interests corrupt public health research*. Cambridge, MA: Harvard University Press.

McGregor, D. (1994). Indigenous knowledge in Canada: Shifting paradigms and the influence of first nations advocates. In T. Veeman, D. Smith, B. Purdy, F. Salkie, & G. Larkin (Eds.), *Science and practice: Sustaining the boreal forest* (pp. 192–197). Edmonton, Alberta, Canada: Sustainable Forest Management Network.

Oreskes, N., & Conway, E. M. (2010). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming* (1st ed.). New York, NY: Bloomsbury Press.

Selinger, E., & Crease, R. P. (Eds.). (2006). *The philosophy of expertise*. New York, NY: Columbia University Press.

Shrader-Frechette, K. S. (1991). *Risk and rationality: Philosophical foundations for populist reforms*. Berkeley: University of California Press.

Thompson, P. B., & Dean, W. (1996). Competing conceptions of risk. *Risk: Health, Safety and Environment, 7,* 361–384.

Turner, S. (2003). *Liberal democracy 3.0: Civil society in an age of experts*. Thousand Oaks, CA: Sage.

Whyte, K. P., & Crease, R. (2010). Trust, expertise and the philosophy of science. *Synthese, 177*(3), 411–425.

Philosophy of History

What is history? What is involved in historical research and knowledge creation? What is the role for the philosophy of history? We might say that there are basically three relevant activities in the realm of historical thinking. First, there is first-order research into facts about the past based on currently available evidence. Second, there is analysis and explanation of those facts, including methods ranging from ethnography to process tracing to structured comparisons to application of findings from the social sciences. These two parts of historical research are guided by ensembles of historical methods and practices (historiography). And, third, there is philosophical reflection on the logic and limitations of these processes of inquiry and inference. This aspect of the discussion includes the conceptual work that philosophers are well qualified for—posing and answering questions like "What is history?" "What is a cause?" or "What is involved in expressing a meaningful action?"

History and Historical Inquiry

Let us begin by attempting to specify the meaning of the word *history*. Here is a schematic definition: History is the sum total of human actions, thoughts, creations, and institutions, arranged in temporal order. Call this "substantive history." History is social action in time, performed by a specific population at a time. Individuals act, contribute to social institutions, and contribute to change. People had beliefs and modes of behavior in the past. They did various things. Their activities were embedded within, and in turn constituted, social institutions at a variety of levels. Social institutions, structures, and ideologies supervene upon the historical individuals of a time. Institutions may have great depth, breadth, and complexity. Institutions, structures, and ideologies display dynamics of change that derive ultimately from the mentalities and actions of the individuals who inhabit them during a period of time. And both behavior and institutions change over time. "History" is the temporally ordered sum of all these facts.

We are interested in understanding history for a number of reasons.

a. We are interested in knowing how people lived and thought in times and settings very distant from our own. What was it like to be a medieval baker or a beadle or a wife? Understanding this kind of thing has a lot in common with ethnography or interpretive research; historians uncover what they can of the circumstances, actions, and symbols of a group of people, and they try to reconstruct

their mentality and their reasons for acting as they did. Robert Darnton's *The Great Cat Massacre* is a good example.

b. We are interested in the concrete social arrangements and institutions that existed at various points in time. We would like to know how marriage or tax collecting worked in rural Ming China. Understanding this kind of thing requires careful study of existing records that permit inferences about how basic institutions worked. Examples include bodies of law, charters, manorial records, and the like. Marc Bloch illustrates this kind of research in *Feudal Society*.

c. We are interested in the dynamics of change—the reasons for the rise and fall of the Roman Empire, the reasons for a rash of peasant rebellions in Qing dynasty China, or the reasons for the occurrence and characteristics of the Industrial Revolution. Understanding this kind of thing has to do with identifying dynamic causal processes of the sort that the social sciences study—why legislatures tend toward certain kinds of institutions and behaviors, why bureaucracies tend toward rigidity, why people are susceptible to extremism. This kind of question pays attention to both internal reasons for change and external reasons—an internal dynamic toward dynastic instability and an external shock imposed by sudden climate change, for example. Charles Tilly has contributed a great deal on this approach, especially in *The Contentious French*.

d. We are interested in quantitative assembly of historical data—population, economic activity, and other kinds of social data. Understanding this kind of thing requires discovering data sources in the historical records and archives that permit estimation of things like marriage rates, grain prices, or church membership totals and then analyzing and presenting these data in convincing ways using established methods in social science quantitative methodologies. A very good illustration of this kind of research is provided by Bozhong Li in *Agrarian Development in Jiangnan, 1620–1850*.

This suggests a fairly simple logic of historical inquiry and representation:

- Historians discover factual circumstances about conditions of life, action, and thought (mentality) during specific periods.
- Historians identify changes in these conditions from one period to another.
- Historians identify the features of social relationships, institutions, structures, and ideologies during specific periods.
- Historians use "path-tracing" and comparison methodologies to discern how circumstances and actions in one period led to specific outcomes in a later period.
- Historians make use of the findings of the social sciences to identify the social dynamics associated with specific kinds of social institutions, structures, and ways of thinking.

This description leaves out a great deal of what historians spend a lot of time on: formulating narratives that make sense, discovering unexpected causes or outcomes of historical circumstances, finding new perspectives on old historical questions, or just deciphering what is going on in an archival source, for example. The goal here is to strip away those elements of the historian's work in order to highlight the logic of the varieties of factual and explanatory claims that historians make. This description is abstract, of course, but it seems to capture the main elements of historical cognition. An important recent turn in the philosophy of history addresses precisely these features of historical inquiry and writing.

History and Philosophy

So far we have analyzed *history*. What about the *philosophy of history*? What is the work that we can hope to do with a philosophy of history? What is it about history that supports philosophical inquiry? Why is the enterprise of investigating, documenting, and explaining facts about the past amenable to philosophical study? What makes the effort to arrive at knowledge of facts about the past an area of philosophical concern?

There are a couple of matters that are relatively clear. First, the domain of historical knowledge is a familiar subject for philosophical inquiry. Epistemology is the philosophy of knowledge generally. Philosophy of science is the careful analysis of the methods and justification of scientific knowledge. And philosophy of historical knowledge is likewise an epistemic and methodological domain: How do we know about the past? Are there limits to our possible knowledge about the past? What is involved in explaining a historical fact? Answers to these questions and others

in the same vein may be difficult and controversial, but it is clear how they fit into existing conceptions of philosophical inquiry. Analytical philosophers like Patrick Gardiner, Carl Hempel, and William Dray have focused on this set of issues.

Second, there is a serious debate about the relative importance of hermeneutic and causal methods for historical inquiry and explanation. Some historians and philosophers have argued that history concerns human action, which is inherently meaningful; so it is necessary for historians to use methods of interpretation of actors and their products that derive from the hermeneutic tradition of Wilhelm Dilthey and Friedrich Schleiermacher if they are to succeed in their tasks. This set of topics intersects profoundly with the philosophy of action—the philosophical study of the facets of intentional individual thinking, choice, and behavior. Recent analytical philosophy (in the work of Harry Frankfurt, Elisabeth Anscombe, or Michael Bratman) has contributed significantly to these issues in the past few decades. This field is one where there has been productive work in both the Continental and the analytic traditions. Contemporary Continental voices contributing to this set of issues include Paul Ricoeur and Hans-Georg Gadamer.

Third is an area of philosophical investigation that arises in connection with the idea of historical and social causation. Some historians and philosophers acknowledge the role of intentional actors in history but argue that the tools of causal analysis are most valuable for historians: For example, this war occurred because conditions X, Y, and Z existed and instigating event E occurred. This debate raises questions across several philosophical fields, in that it requires that we have a more clearly specified theory of causation that is relevant to the social world.

If the philosophy of history were limited to these sorts of questions, then it would be an amalgam of the philosophy of science, the philosophy of action, and the philosophy of causation. However, some philosophers have felt that substantive history itself raises questions that do not reduce to questions about knowledge of these familiar domains.

In particular, some philosophers and historians have argued that we need a metaphysics of history—an account of the kinds of things, forces, structures, and direction that exist in the realm of history. Early exponents of this view of history include Giambattista Vico and Johann Gottfried von Herder. Perhaps the metaphysics of history can shed light on what kinds of structures and entities travel through history and what kinds of processes and systems propel change in the structures and entities.

For example, Karl Marx wrote that "history is the history of class conflict." This implies that classes are historical objects—they exist in the flow of historical events. Other historians have said things like this: Classical slavery gave way to feudalism, which was followed by capitalism. This formulation presupposes that large social-economic systems—for example, social property systems—exist in history and conform to some set of dynamics. And yet others have tried to carve up world history into a set of more or less distinct civilizations—bodies of values, ideas, identities, and institutions that differ significantly from one another.

We might better say, however, that none of these questions pertains to a metaphysics of history, but rather they pertain to a metaphysics of the social world. History is about change and transformation, but the subject of change is the social structures, cultures, and agents that exist within the social world at a period of time. So classes, social property systems, ideologies, and religions are all social arrangements that change over time. If there is such a thing as a "civilization," this is a fact about society at a certain time and not a fact about the structure of history. History has to do with events and dynamic properties; the social world encompasses everything that is happening at a moment in time. We might consider an analogy with the ocean—"history" is the waves, society is the water.

We might be more inclined to recognize a metaphysics of history if we thought there were enduring temporal structures that could be discovered in substantive history—perhaps the business cycle, the Kondratiev long-wave economic cycle, or the rise and fall of civilizations. In other words, if we thought that events and changes conformed to a higher-level pattern of temporal change, we might want to say that the metatemporal pattern is a metaphysical characteristic of history. This was the inspiration of 18th- and 19th-century Continental discussions of history. Enlightenment philosophers discerned a steady movement toward a more rational world, as the Marquis de Condorcet did, and G. W. F. Hegel located an underlying rationality within the events of history themselves. But it is more persuasive to argue that these patterns of change too represent nothing more than empirical characteristics of social phenomena over time. They need to be explained using the tools of the social sciences and are not properly the subject of a priori investigation.

This discussion suggests a few conclusions about the possible scope of research in the philosophy of history:

- Philosophical work directed toward elucidating the epistemology and methodology of historical knowledge is straightforward but limited.
- Philosophical reflections on the nature of action and causation are valuable contributions to the philosophy of history.
- It is true that some historical investigation requires "hermeneutic" interpretation of meanings; but these efforts always fall at the level of interpreting the meaning of individual actions and their meaningful creations, not the large expanse of history.
- The dynamics of historical change are properly the subject matter of the social sciences rather than the philosopher of history.
- Philosophical work directed toward elucidating the metaphysical underpinnings of "history" will be disappointing; those foundations exist within the domain of the philosophy of the social sciences rather than the philosophy of history.

Conclusion

This analysis suggests that the primary tasks for the philosophy of history lie within the philosophy of action, the philosophy of causation, and a specialized portion of the philosophy of science. How do we know about the past? How do we understand the actions of historically situated actors? What is involved in making assertions of causality within the play of history? When these sorts of questions are addressed in the writings of talented historians, the results of work in the philosophy of history can be very illuminating.

Daniel Little

See also Action, Philosophical Theory of; *Annales* School; Causation, Philosophical Views of; Causation in the Social Sciences; Causes Versus Reasons in Action Explanation; Explanation, Theories of; Herder's Philosophy of History; Historicism; Holism, in the Social Sciences; Individualism, Methodological; Marxism and Social/Historical Explanation; Narrative in Historical Explanation; *Naturwissenschaften* Versus *Geisteswissenschaften*; Vico's *Scienza Nuova*; World-Systems Analysis

Further Readings

Ankersmit, F. R. (2001). *Historical representation: Cultural memory in the present*. Stanford, CA: Stanford University Press.

Ankersmit, F. R., & Kellner, H. (Eds.). (1995). *A new philosophy of history*. Chicago, IL: University of Chicago Press.

Anscombe, E. (2000). *Intention*. Cambridge, MA: Harvard University Press. (Original work published 1957)

Bloch, M. L. B. (1964). *Feudal society*. Chicago, IL: University of Chicago Press.

Bratman, M. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.

Clark, E. (2004). *History, theory, text: Historians and the linguistic turn*. Cambridge, MA: Harvard University Press.

Condorcet, J.-A.-N. de C. (1979). *Sketch for a historical picture of the progress of the human mind*. Westport, CT: Greenwood.

Danto, A. (1985). *Narration and knowledge: Analytical philosophy of history*. New York, NY: Columbia University Press. (Original work published 1964 as *Analytical Philosophy of History*)

Darnton, R. (1984). *The great cat massacre and other episodes in French cultural history*. New York, NY: Basic Books.

Dilthey, W. (1989). *Introduction to the human sciences* (R. A. Makkreel & F. Rodi, Eds.). Princeton, NJ: Princeton University Press.

Dray, W. (1957). *Laws and explanation in history*. London, England: Oxford University Press.

Frankfurt, H. (1999). *Volition, necessity, and love*. Cambridge, England: Cambridge University Press.

Gadamer, H. G. (1977). *Philosophical hermeneutics* (1st ed.). Berkeley: University of California Press.

Gardiner, P. L. (1952). *The nature of historical explanation*. London, England: Oxford University Press.

Hegel, G. W. F. (1975). *Lectures on the philosophy of world history* (H. B. Nisbet, Trans.). Cambridge, England: Cambridge University Press.

Hempel, C. (1942). The function of general laws in history. In C. Hempel (Ed.), *Aspects of scientific explanation* (pp. 231–243). New York, NY: Free Press.

Herder, J. G. (1968). Reflections on the philosophy of the history of mankind. In F. E. Manuel (Ed.), *Classic European historians*. Chicago, IL: University of Chicago Press. (Original work published 1791)

Little, D. (2010). *New contributions to the philosophy of history*. Dordrecht, Netherlands: Springer.

Marx, K., & Engels, F. (1974). The Communist manifesto. In D. Fernbach (Ed.), *Political writings: Vol. 1. The revolutions of 1848* (pp. 62–98) London, England: Penguin. (Original work published 1848)

Ricoeur, P. (2004). *Memory, history, forgetting*. Chicago, IL: University of Chicago Press.

Schleiermacher, F. (1998). *Hermeneutics and criticism and other writings* (A. Bowie, Ed.; Cambridge Texts in the

History of Philosophy). New York, NY: Cambridge University Press.
Tilly, C. (1986). *The contentious French: Four centuries of popular struggle*. Cambridge, MA: Harvard University Press.
Vico, G. (2002). *The first new science* (L. Pompa, Ed.; Cambridge Texts in the History of Political Thought). New York, NY: Cambridge University Press. (Original work published 1725)

Philosophy of Politics, History of

The entry introduces the foundational questions raised by a philosophical reflection on the nature of politics as a social science and traces their provenance in the history of political philosophy.

Questions

The history of the philosophy of politics may be defined along a spectrum: At one end are theories treating politics as largely derivative of other domains of philosophy or science, and at the other are those stressing its significantly autonomous nature.

Both of these opposing tendencies address fundamental questions about the *nature* of politics. Is there a distinct domain of the political, distinct, for example, from the moral or ethical, from the economic, from the religious? Is politics a universal feature of human society or a mode of interaction that arises only in certain contexts, and is it historically mutable? Should one think about politics in terms of scientific approaches, such as the implications of neuroscience; or in terms of mathematical approaches to instrumental reasoning, such as rational choice theory; or in terms of the underlying economic forces, as Marxism would contend? Or should one approach politics by means of an interpretative understanding of the meanings of a social and political world, as Max Weber would insist?

A related set of questions is what the philosophy of politics is *for*: Is it designed to counsel political agents as such, or can and should it be the detached knowledge of a political scientist? Should it be addressed to the outlining of a utopian ideal or to political life as it is currently practiced? How a philosopher or political theorist answers such questions, whether implicitly or explicitly, sketches out the philosophy of politics that she or he adopts.

History

The polar ends of the spectrum of the philosophy of politics were influentially modeled among the ancient Greeks. Indeed, it has been argued that the ancient Greeks themselves invented the domain of the political and the idea of politics in their practice of a more or less widely shared civic role of deliberation and execution of policies for the security and flourishing of the community. One line of philosophical interpretation of this domain is exemplified by Aristotle's integral connection of ethics and politics, both aiming at the flourishing of the individual, understood as possible only as a full participant in a political community. Aristotle held ethics and political science to be forms of practical wisdom dealing distinctively with choices about matters that could be otherwise, as opposed to philosophical contemplation, which deals with truths that are independent of human decision. He articulated this view in criticism of the prior expression of roughly the polar opposite view on our spectrum in Plato's *Republic*, which treated political expertise as depending crucially (though not exclusively) on philosophical knowledge of the nature of reality. Thereafter, from the Roman and medieval periods through the 20th century, Aristotelian approaches to the philosophy of politics (and also Epicurean ones) have stressed the relative autonomy and distinctiveness of politics, while Platonic approaches (and also Stoic ones) have stressed its dependence on a wider and higher philosophical framework.

Of course, the ensuing centuries have brought innovation as well as recurrent recourse to the Platonic and Aristotelian extremes. On the one hand, both of these approaches came under fire for being excessively idealistic about politics and its ties to ethics and the moral virtues, with Niccolò Machiavelli in the early 16th century taking princely politics to be neither a philosophical deduction nor an optimistic domain of flourishing but rather a bold and morally risky venture or art (though in republics it remained a form of action for shared liberty). On the other hand, the aspiration to ground politics in knowledge shifted from the domain of philosophy to the domain of science, with Thomas Hobbes in the 17th century deriving laws of reason as basic theorems from a scientific account of the human

condition (though Hobbes was also aware of the role of rhetoric in persuading people to adopt such a rational outlook). Thus, Machiavelli and Hobbes came to stand for two further types of approach to the philosophy of politics, the former treating politics as an art of executive action, the latter as a would-be science of justification.

The history of philosophy of politics in the 20th century can be broadly schematized along the lines of these forebears, notwithstanding elements of novelty. Some thinkers questioned the autonomy of politics, elevating economics (as did Marxists), psychology (as did Freudians), or even metaphysics (as did the followers of Martin Heidegger) to a higher explanatory plane. Others, including Martin Heidegger's sometime student and lover Hannah Arendt, reasserted that autonomy, defending the political as a threatened space of disclosure of individual distinction in terms influenced by Aristotle and ancient Greek practice. Still others, such as Carl Schmitt, a self-described (but arguably inauthentic) follower of Hobbes, insisted that politics was not the space of deliberation but rather the space of a unique, nonderivative domain of existential opposition between friend and enemy. These philosophers agreed at least that politics was not a domain of instrumental reason and that it could not be studied by descriptive social science. Yet a descriptive and even predictive study of instrumental political reasoning was strongly defended by partisans of the new approaches of behavioral and, later, rational choice political science.

Today, the philosophy of politics remains a contested subject, with political science departments divided among interpretative and historical approaches, rational choice modeling, and scholars of the historical development of the philosophy of politics itself. An alternative line of thought is the revival of republicanism in contemporary political theory.

Melissa Lane

See also Evolutionary Political Science; Hobbes's Philosophical Method: Nature–Man–Society; Machiavelli's Art of Politics; Neuroscience and Politics; Rational Choice and Political Science; Utopianism; Weber's *Verstehende* Approach

Further Readings

Arendt, H. (1958). *The human condition.* Chicago, IL: University of Chicago Press.

Kaldis, B. (1999). Aristotle and Kant on contemplation, ethics and politics. *Archiv für Rechts- und Sozialphilosophie, 85*(4), 540–562.

Kelly, D. (2003). *The state of the political: Conceptions of politics and the state in the thought of Max Weber, Carl Schmitt, and Franz Neumann.* Oxford, England: Oxford University Press.

Lane, M. (1998). *Method and politics in Plato's statesman.* Cambridge, England: Cambridge University Press.

Lane, M. (2001). *Plato's progeny: How Plato and Socrates still captivate the modern mind.* London, England: Duckworth.

Meier, C. (1990). *The Greek discovery of politics* (D. McClintock, Trans.). Cambridge, MA: Harvard University Press. (Original work published 1980)

Pettit, P. (1997). *Republicanism: A theory of freedom and government.* Oxford, England: Clarendon.

Viroli, M. (1992). *From politics to reason of state: The acquisition and transformation of the language of politics, 1250–1600.* Cambridge, England: Cambridge University Press.

Philosophy of Sociology, History of

Philosophy and sociology have a complex and intertwined historical relationship. Early figures in sociology, such as Auguste Comte (1798–1857), the inventor of the term, also figure in the history of philosophy of science as the inventors of "positivism." In this case, the philosophy served as a methodological warrant for the claim that the "sociological" theories of Comte were "scientific," and the problem of justifying particular forms of sociology as "science" was a persistent theme in both philosophical writing on social science and the field of sociology itself. Nevertheless, other philosophical sources, notably Immanuel Kant, played and continue to play an important role both in the problematic relations between the fields and as a philosophical source for sociology.

The basic problem posed by sociology for the philosophy of science is this: There seems to be no reason in principle why the social world should not be understandable scientifically. But finding the right analogs to existing science proved difficult. Comte had to rethink the notion of scientific law and scientific development itself in order to construct such an account, and his result combined both: Law was no more than prediction, and the scientific character

of sociology was confirmed by the possession of a law, the three-stage law of the development of the sciences through the stages of theological, metaphysical, and positive, which itself predicted the development of sociology as a science.

This clever solution had many consequences, not the least for the anti-metaphysical program of logical positivism itself. But it also spurred John Stuart Mill (1806–1873), in Book Six of *A System of Logic*, to construct a different account of the place of sociology in relation to other sciences, as well as a reconsideration of the notion of scientific law as it might apply to the various social sciences. Mill recognized that problems of casual complexity, the probabilistic character of laws, the problems of combining laws, and the problems of relating different kinds of social-scientific reasoning, including his associational psychology and deductive economic theory, meant that at best social science could produce approximations to laws.

Even finding approximations to laws proved difficult, so core Millian ideas were replaced and revised in various directions. Émile Durkheim (1858–1917) tried to construct laws governing the probabilities of suicide for different categories. Max Weber (1864–1920) drew on an alternative account of probability, which he used to warrant causal attribution in ordinary-language terms in the absence of laws. The mathematician Karl Pearson's correlational methods were used by American sociologists, notably F. H. Giddings and his students, following Pearson and the physicist Ernst Mach's slogans: Mach's that laws were no more than descriptions and Pearson's that supposed laws even in physics were idealized correlations. Other forms of sociology relied on analogies with organisms or with organic evolution, appealing to teleological or functional explanation.

These ideas of sociology as a science produced an "idealist" response that drew on various sources and took various forms. One was the idea of understanding, or *Verstehen*, and the idea that the human realm required this kind of knowledge rather than law. Another was an application of the Neo-Kantian idea of science, which took a science to be a hierarchically ordered realm of concepts that constituted a domain of fact. Scientific understanding, on this account, was primarily conceptual understanding of a world, in this case the social world, that is already conceptualized.

The Neo-Kantian approach was applied in early German sociology from the early work of Georg Simmel (1858–1918), who asked the self-consciously Kantian question "How is society possible?" and constructed a system based on the concept of the social. The autonomy of sociology in this approach depended on the existence of such a distinct ordering concept. German sociology became a series of constructions of this kind, based on different concepts of the social. In the 1930s, the American sociologist Talcott Parsons (1902–1979) produced a famous version that proved influential in the American context, and later the world, but he gave a meta-theoretical justification of it that linked it to approximations in physical science.

The relationship between sociology and philosophy becomes confused at this point, the point at which logical positivism begins to dominate American philosophy. Attempts to make sense of Parsons's ideas of "theory" in terms of logical positivism, or indeed any notion of theory that was explanatory, were unavailing. Philosophers like Carl Hempel and Thomas Nagel made an extended effort to explicate sociological thinking in the 1950s, primarily in terms of the functionalist aspects of Parsons's and also Robert Merton's (1910–2003) writings. But Parsons eventually denied that he was a functionalist. The English philosopher Max Black analyzed Parsons's writings at length and concluded that they were essentially devoid of meaning or expressed tautologies or truisms.

Sociologists ransacked logical positivism for hints at how to make their theories scientific. But the various borrowings from logical positivism never meshed very well with the actual forms of statistical analysis normally practiced in sociology. Instead, a practical convergence occurred with variants on the casual modeling methods of Herbert Simon, which were taken up eagerly by sociologists in the 1960s and 1970s, initially in the form of path analysis. Later, these became important topics in philosophy, in the work of Clark Glymour, James Woodward, and Judea Pearl.

Peter Winch's *The Idea of a Social Science*, which appeared in 1958, provided a "Wittgensteinian" critique of sociology and suggested that it was better understood as a form of conceptual analysis, like philosophy, but with the aim of understanding the concepts of a particular society or understanding that social action was to be placed within language-games and understood in terms of rule following. Karl Popper, in *The Poverty of Historicism*, suggested that explanations of action in terms of what

he called the logic of the situation were the appropriate form of social explanation and required only considerations of instrumental rationality. The latter prefigured rational choice explanation and restated Weber; the former reinforced the idea that the only proper explanations of human action were in terms of the concepts of the agent. Little of this survived the criticisms of Donald Davidson in articles like "The Very Idea of a Conceptual Scheme."

Stephen Turner

See also Causation in the Social Sciences; Causes Versus Reasons in Action Explanation; Durkheim's Philosophy of Social Science; Language-Games and Forms of Life; Logical Positivism/Logical Empiricism; Neo-Kantianism; Paradigms of Social Science; Positivism, History of; Rule Following; Simmel's Philosophy of Society; Situational Logic; Structural Functionalism, in Social Theory; Weber and Social Science: Methodological Precepts; Weber's *Verstehende* Approach

Further Readings

Manicas, P. (1987). *History and philosophy of the social sciences*. Oxford, England: Blackwell.

Nagel, E. (1961). Methodological problems of the social sciences. In *The structure of science: Problems in the logic of scientific explanation* (pp. 447–546). New York, NY: Harcourt, Brace & World.

Turner, S. (1986). *The search for a methodology of social science*. Dordrecht, Netherlands: Reidel.

Plural Subjects

This entry reviews Margaret Gilbert's theory of *plural subjects*, draws parallels to and differences from other sociological traditions, and goes on to provide a critical examination of its various aspects.

The notion of "plural subject" was coined by Margaret Gilbert in *On Social Fact*. It has played an increasing role in the contemporary philosophy of social science and in the social sciences themselves. The notion of the plural subject is at the core of Gilbert's philosophy, along with a closely related notion, that of *joint commitment*. This conceptual link became clear in Gilbert's (2006) more recent definition: "A and B (. . .) constitute a plural subject (. . .) if and only if they are jointly committed to doing something as a body" (p. 144).

The main interest of the notion of plural subject is that it consists of an attempt to give a more analytical and simpler form to "holistic" intuitions that have been formulated either by sociologists such as Émile Durkheim or by idealist philosophers such as G. W. F. Hegel. Gilbert's challenge is to try to reconstruct these intuitions within the limits of methodological individualism (in a version much closer to Georg Simmel's than to Marx Weber's) and on contractarian bases, which, at first sight, seems paradoxical.

Like Simmel and more clearly than Weber, Gilbert assumes that groups are not "reducible" to individuals but are nevertheless "constructed" by interactions or relations between human individuals. In this way, the notion of plural subject is clearly framed in a constructivist view of social life. However, unlike interactionist sociologists (e.g., Erving Goffman) or relationist social scientists (e.g., James Coleman), Gilbert does not think that common interactions or relations (e.g., trust) are sufficient as bases of social life. Moreover, unlike David Hume and Adam Smith or, more recently, Michael Bacharach, she does not think that identification as a very specific relation would fill the gap. On the contrary, contracts—in the sense of *actual* contracts (and not in the sense of hypothetical contracts, as in Thomas Hobbes, Jean-Jacques Rousseau, or John Rawls)—are required.

Most of these "contracts," Gilbert argues, are *tacit* or implicit, unlike formal contracts but like informal agreements. They are not necessarily based on a cost–benefit calculation from the viewpoint of participants' self-interest, and it is not even necessary that the association forged by these agreements be primarily self-interested either. What is necessary, in any case, is that there be a reciprocal *commitment* of each participant to the common goal (whatever it is) but, furthermore, that this reciprocal commitment be a *joint* commitment. This means that the realization of the common goal (e.g., to play a symphony, to make a revolution, to increase the income of a company) has to be the priority, and each participant in the collective action has to act as a mere member of the social body ("as a body": Gilbert's own phrase, used frequently by her). This is a crucial point that is not expressed by the simpler notion of reciprocal commitment. It should be mentioned that the above definition of plural subjects is redundant (i.e., "as a body" means nothing more than "jointly") yet not circular as several scholars have argued.

Gilbert avoids using both the notion of collective intentionality, much debated since John Searle,

and Raimo Tuomela's more idiosyncratic notion of "we-mode." It seems obvious, however, that the concept of joint commitment requires notions similar to these. Speaking, for example, of a "plural subject collective intentionality" would be relevant. Unfortunately, neither Searle nor Tuomela have used the idea of commitment, an element inherent in the proper understanding of collective intentionality according to Gilbert. Michael Bratman tackled this issue carefully, arguing that, although Gilbert might be right generally speaking, there is however a more basic notion of collective intentionality (or, as he says, "shared intentionality")—or of plural subjecthood—that does not require the idea of commitment. Bratman urges that we should consider Hume's well-known example of two people in a boat rowing together without ever having given promises to each other: It would be natural to take these rowers as having a shared intention to row the boat together. In these cases, mere mutual adjustment is required. One could argue, however, that promise or, at least, joint commitment is implicit (Gilbert argues that joint commitment is not based on promises but that, in fact, the opposite is true) and that if one rower did not attempt to adapt his rhythm to that of the other rowers, they would feel entitled to reproach him.

The notion of plural subject is *normative* in the sense that it involves the normative notion of joint commitment. However, joint commitments are neutral ethically speaking, and one might have to violate joint commitments to fulfill moral commitments (Nazi leaders were probably jointly committed and constituted a plural subject). In this respect, Gilbert's notion of joint commitment and Amartya Sen's notion of commitment differ since, according to Sen, commitment means moral duty. On the other hand, if one views political obligations as ethically neutral, the notion of joint commitment can be thought of as the basis of political obligations.

Other reservations might come from the phenomenological method favored by Gilbert. As she regularly refers to subjective experiences (e.g., the feeling of being committed, the feeling of acting as a body), it may be argued that these mental states might not be universal. Psychological experiments should be conducted. However, until now, psychologists have been especially interested in searching for the anticipation of plural subjectivity (involving joint commitments) in children. Primatologists also are interested in similar possible anticipations in apes, although current investigations are not refined enough to provide any satisfying response here. Indeed, the idea of normativity inherent in joint commitment is often forgotten, and the notion of joint commitment is often implicitly replaced by the less demanding notion of shared intentionality.

Alban Bouvier

See also Collective Agents; Collective Intentionality; Commitment; Holism, in the Social Science; Individualism, Methodological; Searle and the Construction of Social Reality; Simmel's Philosophy of Society; Social Contract Theories; Social Facts; Social Ontology, Recent Theories of; Social Rules; We-Mode, Tuomela's Theory of

Further Readings

Bouvier. A. (2004). Individual and collective beliefs in sciences and philosophy: The plural subject and the polyphonic subject accounts. Case studies. *Philosophy of the Social Sciences, 34,* 382–407.

Bratman, M. (1999). *Faces of intention.* Cambridge, England: Cambridge University Press.

Bratman, M. (2007). *Structures of agency.* New York, NY: Oxford University Press.

Gilbert, M. (1989). *On social fact.* Princeton, NJ: Princeton University Press.

Gilbert, M. (2006). *A theory of political obligation.* Oxford, England: Oxford University Press.

Policy Applications of the Social Sciences

The application of social science knowledge to improve public policies is of great interest to philosophers, social scientists, and laypersons. Although "science" as a specialized form of reliable knowledge did not develop until the 19th century, the production and use of diverse forms of policy-relevant knowledge is as old as civilization itself.

Emergence of Policy-Relevant Knowledge

Policy-relevant knowledge emerged at a point in the evolution of human societies where practical knowledge was consciously cultivated, thereby prompting a self-reflective examination of links between knowledge and action. What we now know as the "policy sciences"—namely, policy-oriented social

sciences including political science, sociology, and economics—are a modern attempt to promote practical knowledge, that is, knowledge about what works.

Specialized procedures for analyzing policies arose with the emergence of urban civilization out of scattered tribal and folk societies. The policy sciences followed the changes in social and, above all, political organization that accompanied new production technologies and stable patterns of human settlement. For example, early Mesopotamian legal codes were a response to the growing complexity of fixed urban settlements, where policies were needed to regulate the distribution of commodities and services, organize record keeping, and maintain internal security and external defense. A growing consciousness of relations between knowledge and action fostered the growth of educated strata that specialized in the production of policy-relevant information. These "symbol specialists," as Harold Lasswell called them, were responsible for early forms of policy forecasting. Symbol specialists were expected to foresee crop yields at the onset of the planting season or predict the outcomes of war. Because these early producers of policy-relevant knowledge used mysticism, ritual, and the occult to forecast the future, their methods were unscientific by present-day standards.

Then as now, however, policy-relevant knowledge was ultimately judged according to its success (or failure) in shaping better policies, not simply because particular methods were used to produce it. Even the ancients seemed to know what some contemporary analysts forget—when methods are used for ritualistic purification, political persuasion, or symbolic legitimation, analysts and their clients eventually must face the decisive test of performance, namely, the pragmatic test of whether the use of specialized procedures produces better policies.

Over the centuries, there have been preeminent individual producers of specialized knowledge, for example, Aristotle in his role as tutor to Alexander the Great. However, entire classes of educated persons would later influence policy making in Europe and Asia. In the Middle Ages, the gradual expansion of urban civilization brought with it an occupational structure that facilitated the development of specialized knowledge. Princes and kings recruited policy specialists to provide advice and technical assistance in areas where rulers were least able to make effective decisions: finance, war, and law.

The age of the Industrial Revolution was also that of the Enlightenment, a period in which a belief in human *progress* through science and technology became an ever more dominant theme among policymakers and their advisers. The development and testing of scientific theories of nature and society gradually came to be seen as the only objective means for understanding and solving social problems. For the first time, policy-relevant knowledge was produced according to the canons of empiricism and the scientific method.

The 19th-Century Transformation

In 19th-century Europe, producers of policy-relevant knowledge began to base their work on the systematic recording of *empirical data*. Earlier, for several thousand years, there was an essential continuity in methods for investigating and solving social, economic, and political human problems. If evidence for a particular point of view was provided, it was typically based on appeals to religious authority, ritual, or philosophical doctrine. What was new in the 19th century was a basic change in the procedures used to understand society and its problems, a change reflected in the growth of empirical, quantitative, and policy-oriented research.

The first censuses were conducted in the United States (1790) and England (1801). It was at this time that statistics ("state arithmetic") and demography began to develop as specialized fields. The Manchester and London Statistical Societies, established in the 1830s, helped shape a new orientation toward policy-relevant knowledge. Organized by bankers, industrialists, and scholars, the societies sought to replace traditional ways of thinking about social problems with empirical analyses of the effects of urbanization and unemployment on the lives of workers and their families. In the Manchester Society, an enthusiasm for quantification was coupled with a commitment to social reform or "progress of social improvement in the manufacturing population." The London Society, under the influence of Thomas Malthus (1766–1834) and other academics, took a more disinterested approach:

> The Statistical Society will consider it to be the first and most essential rule of its conduct to exclude carefully all opinions from its transactions and publications—to confine its attention rigorously to facts—and, as far as it may be found possible, to

> facts which can be stated numerically and arranged in tables. (Statistical Society of London, 1838, p. 70)

The London and Manchester societies used questionnaires to carry out studies, and paid "agents" were the counterpart of today's professional interviewer. There were similar developments in France, Germany, and the Netherlands.

A preeminent contributor to the methodology of social statistics and survey research was Adolphe Quetelet (1796–1874), a Belgian mathematician and astronomer, who was the major scientific adviser to the Dutch and Belgian governments. In contemporary texts on designing social and economic surveys, there are many topics that were addressed by Quetelet: questionnaire design; data collection, analysis, and interpretation; data organization and storage; and identification of the conditions under which data are collected. In the same period, Frederic Le Play (1806–1882) wrote *Les Ouvriers Europeans* (The European workers), a detailed empirical investigation of the family income and expenditures of European workers in several countries. In Germany, Ernst Engel (1821–1896) sought to derive laws of "social economics" from empirical data expressed in statistical form.

In England, the work of Henry Mayhew and Charles Booth, who studied the life and employment conditions of the urban poor in natural (what we now call "field") settings, is representative of the new empirical approach to the study of social problems. Mayhew's *London Labour and the London Poor* (1851) described the lives of the laborers, peddlers, performers, and prostitutes who constituted London's urban underclass. In writing *Life and Labour of the People in London* (1891–1903), Booth employed school inspectors as key informants. Using what we now call participant observation, Booth lived among the urban poor, gaining firsthand experience of actual living conditions. A member of the Royal Commission on the Poor Law, he was an important influence on the revision of policies on old-age pensions. Booth's work also served as something of a model for policy-oriented research in the United States, including the *Hull House Maps and Papers* (1895) and W. E. B. Dubois's *The Philadelphia Negro* (1899), both of which sought to document the scope and severity of poverty in urban areas.

Dominant social groups valued policy-oriented research as a means to achieve political and administrative control. In the sphere of factory production, for example, the political organization of work preceded scientific and technological developments that later culminated in efficiency-enhancing machinery and the specialization of tasks. Methods of empirical, quantitative, and policy-oriented analysis were a product of the recognition by bankers, industrialists, politicians, and the Victorian middle class that older methods for understanding the natural and social world were no longer adequate. The key questions of the day were practical and political: How much did members of the urban proletariat need to earn to maintain themselves and their families? How much did they have to earn before there was a taxable surplus? How much did they have to save from their earnings to pay for medical treatment and education? How much should capitalist owners and the state invest in day care facilities so that mothers might put in an effective day's work? How much investment in public-works projects—sanitation, sewage, housing, and roads—was required to maintain adequate public health standards, not only to maintain a productive workforce but also to protect the middle and upper classes from infectious diseases cultivated in urban slums?

The 20th Century

An important feature of the 20th century, as compared with the 19th, is the *institutionalization* of the social sciences and social professions. Policy-oriented social scientists played an active role in the administration of Woodrow Wilson, particularly during World War I. Later, under President Herbert Hoover, social scientists carried out major social surveys, *Recent Economic Trends* and *Recent Social Trends*. The largest influx of social scientists into government came, however, with Franklin Roosevelt's New Deal. Social scientists staffed the numerous new agencies established during the Roosevelt administration. During World War II, military and civilian agencies used social scientists to investigate problems of national security, social welfare, war production, pricing, and rationing. The activities of agencies such as the Office of Strategic Services were continued after the war in the form of the Research and Development Board (subsequently RAND) and numerous other policy think tanks.

The landmark work in the development of the policy-oriented social sciences was published in

1951 by the political scientists Daniel Lerner and Harold D. Lasswell. *The Policy Sciences: Recent Developments in Scope and Method* was the first systematic effort to develop an explicit policy orientation within the social sciences. The "policy sciences" were not confined to testing theories; they also had a fundamentally practical orientation. Their purpose was not only to provide a basis for making efficient decisions but also to provide knowledge relevant to the practice of democracy and the realization of democratic values.

Specific methods and techniques of policy analysis grew out of engineering, operations research, and applied mathematics. The idea of "analysis" came to be associated with the decomposition of problems into smaller components; for example, problems of national defense were decomposed into discrete policy alternatives (e.g., nuclear warheads, manned bombers, and conventional ground troops), the benefits and costs of which could be estimated.

This turn toward a more technical perspective was accompanied by the growing influence of nonprofit research organizations ("think tanks") such as the RAND Corporation, which fostered the spread of cost-effectiveness analysis and related techniques to academics as well as policymakers. The idea was to "purchase" national security in the most efficient manner. It is noteworthy that as much as 90% of all research funded by governments, nonprofit organizations, and the private sector is applied research on practical problems.

By the 1970s, many social science disciplines had established institutions expressly committed to policy-oriented research. These include the Policy Studies Organization (Political Science), the Society for the Study of Social Problems (Sociology), and the Society for the Psychological Study of Social Issues (Psychology). In the 1980s, the process of institutionalizing policy-oriented social science was carried a step further by the creation of multidisciplinary professional associations, such as the Association for Public Policy and Management, which holds annual research conferences and publishes a journal of record, the *Journal of Policy Analysis and Management*. The new journal brought a more technical focus than mainstream policy journals, including *Policy Sciences*, the *Policy Studies Journal*, and *Policy Studies Review*. In addition to the mainstream journals were hundreds of others that focused on specific issues involving health, welfare, education, criminal justice, education, science and technology, and other areas.

The 21st Century

In the first decade of the 21st century, there was an increasing recognition that the complexity of problems faced by governments requires the systematic use of the social sciences to develop policies and assess their consequences. The call for *evidence-based policy making* in the United Kingdom, the United States, and the European Union is a response to this complexity; it is also a recognition that ideological, religious, and political influences—usually hidden and lacking in transparency—have exerted a harmful effect on policy making in areas ranging from health, education, and welfare to national security and the environment. Evidence-based policy making in the United Kingdom and the European Union takes several forms, including *regulatory impact assessment*, which refers to the use of social science analyses to examine the benefits, costs, risks, and consequences of policies before they are adopted. In the United States, evidence-based policy has been promoted by leading program evaluators and policy analysts, who founded the Coalition for Evidence-Based Policy of the Council for Excellence in Government. More than 40 countries participate in the Campbell Collaboration, a professional and scientific organization that promotes evidence-based policy. The evidence-based policy movement testifies to the continuing commitment to policy applications of the social sciences.

William N. Dunn

See also Econometrics: Methodological Issues; Evidence-Based Policy; Experimenting Society, The; Experiments in Social Science; Philosophy of Expertise

Further Readings

Dunn, W. N. (2012). *Public policy analysis* (5th ed.). Upper Saddle River, NJ: Pearson.

Glazer, N. (1959). The rise of social research in Europe. In D. Lerner (Ed.), *The human meaning of the social sciences* (pp. 43–72). New York, NY: World Publishing.

Lerner, D., & Lasswell, H. D. (Eds.). (1951). *The policy sciences: Recent developments in scope and method.* Stanford, CA: Stanford University Press.

Statistical Society of London. (1838, April–August). Transactions of the Statistical Society of London, Vol. 1, Pt. 1. *London and Westminster Review, 7&29*, 45–72.

Political Psychology

Political psychology is an interdisciplinary field concerned with research questions that lie at the intersection of psychology and political science. Lacking one widely accepted definition, political psychology might be conceptualized as the study of "how people think about politics and how this cognitive process affects their political behavior" (Monroe et al., 2009, p. 878). We can define political psychology broadly as the study of how patterns of political thinking, feeling, and identity interact to influence political choice and political behavior.

Historical Overview

Political psychology has a long, if informal, tradition. Thinkers from the Greek philosophers to Confucius were fascinated by problems of politics in relation to human nature. Greek mythology and fables captured concepts such as sour grapes. Later political theorists such as Niccoló Machiavelli emphasized the importance of appearance for a political leader. Historians such as Alexis De Tocqueville drew attention to the political importance of psychological phenomena such as the revolution of rising expectations or the feeling of relative deprivation.

The origins of political psychology as a modern scientific field, however, tracks to early- to mid-20th-century scholars such as the German philosopher Theodor Adorno (one of the founders of the Frankfurt school), the American psychologist Gordon Allport, and the American political scientist Harold Lasswell, whose work blended psychology with political behavior to understand the outbreak of wars, political movements, and mass violence. Political psychology became a separate field only in the mid 20th century, in response to dissatisfactions with the behavioral movement. As this new field developed, it contained research that clusters into many related topics: the nature and acquisition of political belief systems, the psychology of leadership and decision making, the cognitive processes involved in warfare, group violence and terrorism, the individual as political actor, political movements, political alignments and structures, intergroup relations, political processes, foreign policy and group think, government and self-esteem, and identity and group conflict.

The first handbook of political psychology, edited by Jeanne Knutson, suggested that when constructing work in political psychology, scholars need to focus attention on the interaction of political and psychological phenomena, to ask how responsive and relevant political psychology is to societal problems and whether context makes a difference for the problems being analyzed. Furthermore, we need to emphasize the process as well as the outcome and to be tolerant of multiple methods for gathering and analyzing data. As a contemporary definition, then, political psychology might be appropriately considered both an approach and a perspective on politics and political life, a way to understand politics that emphasizes the workings of the human mind as it influences human political behavior.

Political psychology itself exists at the nexus of political science and psychology, with scholars of political psychology drawn chiefly from these two disciplines but with important contributions from sociologists, psychiatrists, and scholars of communication, education, and other disciplines that enrich the field. As a field, political psychology utilizes multiple methodological techniques, from surveys and experiments to narratives, interviews, and content analyses. Recent work from neuroscience and biology has enriched and broadened the methodological field.

The Discipline

In contrast to its deep historical roots, solidifying political psychology as an independent discipline began comparatively recently. The field's international organization, the International Society of Political Psychology, was founded in 1978, and the first issue of its journal, *Political Psychology*, appeared in 1979. Less than a dozen graduate programs currently offer degrees or formal certificates in political psychology, though considerably more offer graduate-level courses. The field differs in emphasis, with American schools reflecting a more behavioral approach than European and Australian schools.

Empirical investigation into the nature of political psychology as a discipline uncovered a number of broad topics on which political psychologists initially

focused; chief among these are political beliefs and ideology, personality, cognition, and international relations/intergroup conflict. Scholars such as Lasswell located political psychology's roots in pioneering work on psychopathology and politics; Adorno and his colleagues highlighted the authoritarian personality; and Robert Lane emphasized the origins of political belief systems, especially research on values/ideology and on personality, including work on the psychological correlates of political belief systems. James Sidanius and Felicia Pratto employed work on politically relevant personality traits to develop social dominance theory, and scholars such as Rose McDermott (also a contributor to this encyclopedia) drew on recent work in neuroscience to suggest possible biological bases for political attitudes and the effects of gender differences on decision making. Research by George Lakoff on cognition encompassed work on perception, memory, decision making, and other mental processes as they relate to political activity; it was expanded by Milton Lodge and Kathleen McGraw to explain vote choice and candidate evaluation. Robert Jervis's research in international relations and intergroup conflict encompasses studies of foreign policy decision making. Ervin Staub, Janusz Reykowski, and Kristen Monroe described the psychology of war and genocide, and Gerald Post used psychoanalytic models to explain terrorism. Henri Tajfel and John Turner developed social identity theory and self-categorization theory, which revolutionized our understanding of the psychological roots of the prejudice and discrimination that drive group conflict. Research on the psychology of leaders is a perennial favored topic of political psychologists, from Alexander and Juliette George's classic psychobiography of Woodrow Wilson to more contemporary analyses of the personalities, cognitive traits, and leadership styles of leaders from Wilson and Bill Clinton to Saddam Hussein done by scholars such as David Winter, Rose McDermott, and Gerald Post.

A recent focus of study has been the role of emotion in political decision making and behavior, as found in work by George Marcus. The field itself continues to evolve, with new and exciting work being done on a wide range of topics, such as how migration and multiculturalism influence politics, including consideration of enmity among groups, and how this relates to prejudice, civic and political participation, nationalism, and right-wing extremism. The increasing influence of religion—especially fundamentalism—has spurred exciting work on group identity and political violence and how this relates to religion and civic inclusion/participation. The use of narrative as a tool and how it can yield insight into the workings of the human mind is an especially rich area, as qualitative interpretive analysis is being wedded with sophisticated computer programs from computer science and linguistics to provide insight into the cognitive processes underlying political life.

Kristen Renwick Monroe and Bridgette Portman

See also Affective Intelligence in the Social Sciences; Behavioralism in Political Science; Evolutionary Political Science; Machiavelli's Art of Politics; Neuroscience and Politics; Rational Choice and Political Science

Further Readings

Jervis, R. (1976). *Perception and misperception in international politics*. Princeton, NJ: Princeton University Press.

Knutson, J. N. (1973). *Handbook of political psychology*. Oxford, England: Jossey-Bass.

Lane, R. E. (1962). *Political ideology: Why the American common man believes what he does*. New York, NY: Free Press.

Lasswell, H. (1930). *Psychopathology and politics*. Chicago, IL: University of Chicago Press.

Monroe, K. R. (Ed.). (2002). *Political psychology*. Mahwah, NJ: Lawrence Erlbaum.

Monroe, K. R., Chiu, W., Martin, A., & Portman, B. (2009). What is political psychology? *Perspectives on Politics, 7*(4), 859–882.

Sears, D. O., Huddy, L., & Jervis, R. (Eds.). (2003). *Oxford handbook of political psychology*. New York, NY: Oxford University Press.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks/Cole.

Popper's Philosophy of Science

Traditional philosophies of science going back to Francis Bacon in the 17th century presented science as natural and infallible. It was believed that careful attention to the inerrant methods of science would result in proven truth. This promise clashed with

well-known facts: Science emerged late in human history, against a background of diverse superstition and ignorance; hence, how could it be natural? Its teachings changed over time, and scientists were far from united on all issues. So, far from being infallible, it was neither unified nor stable. Bacon's response was that pride and sloppy mental habits explained our late achievement of science, and their persistence explained the lack of consensus. The Austro-British philosopher Karl Popper (1902–1994) rethought the whole issue by paying close attention to the greatest scientific revolutionary of the 20th century: Albert Einstein. Einstein said that science was the work of the imagination in the service of explaining both the facts and the success and failure of previous scientific explanations of the facts. Popper's key move was to treat the diversity and disputatiousness of science not as a vice but as a virtue: It was that which enabled scientific objectivity and the pursuit of truth. Science required the social discipline of institutions because its critical method did not come naturally, and dispute and argument were needed to test its claims.

Popper's General View of Scientific Method

Popper offered a social rather than a natural account of science as driven by imagination and checked by intersubjective criticism (especially experimental). Science emerges from certain sorts of disciplined and institutionalized social interactions in which participants sometimes hit on good ideas, most of which are sooner or later shown to be false. Science was far from infallible; Popper viewed its fallibility as part and parcel of human fallibility.

Popper's philosophy of science described the logical situation in science as follows. The traditional philosophical view of science as proven truth totally depended upon solving the problem of induction. The problem of induction in Popper's formulation reads thus: How is (theoretical) learning from (particular) experience possible? The 18th-century Scottish philosopher David Hume had shown that solving the problem of induction was logically impossible—hence his ensuing skepticism about the possibility of inductive logic. Hume mistrusted his own results and suggested that we carry on as though nothing were amiss. Einstein said that he had learned from Hume that we can both admire our predecessors and try to improve upon them. Popper elaborated on this significantly and argued that deductive logic, which was not subject to Hume's strictures on induction, was sufficient to account for science, provided we give up all claims to its being infallibly proven and, instead, rest content with the progress that the elimination of error provides. Popper's alternative approach pictured scientists conjecturing theoretical solutions to problems (in well-defined problem-situations) formulated so that testable consequences can be deduced from them. Testing these consequences by observation and utilizing deductive logic, we may be led to contradiction between theory and observation report. In such a case, guessing that our deductions have been sound, and given that the observable consequences contradict our theories or conjectures (premises), we can infer, by deductive logic, that the premises (as a body) are false. The next challenge is to see if the source of error can be narrowed down and isolated. Each stage is conjectural, including all the deductive inferences we make.

The deductive logic of scientific explanation consisted, then, of a set of premises that included theories—that is, general theoretical conjectures—plus particular conjectures about the entities or processes under discussion (initial conditions). From these premises, statements were deduced that describe what is to be expected (predictions). If these latter are simple and precise enough to undergo empirical testing, then that should be undertaken. If the test result is that what is observed conforms to what was predicted, then we learn only that this test did not refute the package of premises. If what is observed contradicts what was predicted, then we can conjecture various possible sources of error, including human error, as well as the possibility that some statements of theory or of initial conditions, or of both, are mistaken. Errors bring our theories into contact with the world. By discovering that they are mistaken, we learn that the world is not as we were conjecturing it to be but otherwise. This does not tell us how the world is but, rather, how it is not. Our theories have bumped into reality.

Criticism and Rebuttals

Some of Popper's critics accuse him of being a thoroughgoing skeptic, a Bad Thing for many philosophers. Others said that deductions rested on induction and so succumb to Hume's criticism. On the contrary, Popper replied, theoretical premises are conjectured, not inferred. He reaffirmed

Bacon's assertion that there is a logical asymmetry between falsification and verification: No amount of evidence allows an inductive inference to a generalization; but a single conflicting instance calls a generalization into question. Nonetheless, such moves remain conjectural, not unerring. Popper thus offered no escape from the reality of human fallibility but offered, instead, comfort in the capacity error offers us to widen our horizons. Scientific success was a matter of imagination, guesses, and tests and not of accumulated evidence toward proof. Pressed to say why science deserved credence, he responded that it did not and that all we knew about severely tested theories was that we had done our best to test them. He did once try to devise a quantitative theory of approximation toward the truth, but his effort was quickly shot down, and he reverted to the qualitative one.

Applications to Social Science

Before World War II, few philosophers and scientists were familiar with Popper's breakthrough. The exceptions included two economists, Terence Hutchison and Friedrich Hayek. Events in Europe in the 1930s sent Popper into permanent exile from his native Austria. He turned his attention to problems of society and politics and discovered there important connections with scientific method. Politics has always been understood as matters for debate and compromise, but there was no coherent methodology for their pursuit. Methodology was infallibilistic; and political thinking was utopian. Popper argued that if science was fallible, then so of course was thinking about society and politics. Utopianism offers solutions to all problems. It thus assumes that we know (or could know) even the unintended outcomes of all actions and find ways to avoid the undesirable ones. It is not hard to show this assumption to be mistaken. Instead of trying to establish politics on a utopian vision, Popper developed a minimal account of democratic and rational politics that closely followed his fallibilist account of science.

Paralleling his emphasis on testability that he connects to falsifiability and falsification, he characterized a free society as one where the ruled could replace the rulers without violence. He insisted that bad government was a much more urgent and serious problem than good government. Bad government played havoc with people's lives; good government attempted to enhance lives. The reduction of suffering, he taught, is more morally urgent than the betterment of life's chances. He found ranged against this simple and intuitive view the majority of philosophers and other intellectuals, who, he claimed, have a tendency to worship power and hence rationalize it. High on his indictment sheet of enemies of freedom were Plato, Aristotle, Hegel, Marx, and many lesser figures.

Popper passionately advocated the open society. He characterized it as one that enabled political participation of all adults; minimized restrictions on individuals, including allowing free exit; encouraged criticism and debate about public policy; and engaged in a good deal of state intervention in order to protect weaker and disadvantaged citizens from avoidable suffering. Such a social order needed social science in order to document social conditions and the effects of government policies. Such social science needed to be grounded in moral concern and responsibility for the alleviation of avoidable suffering. It needed to be conducted with a sense of urgency, since avoidable suffering was intolerable. Otherwise, sciences of society, though similar to natural science in comprising conjectures and refutations in a framework that encourages criticism, differ from it in having to take account not just of morality but also of rationality, that is, people's capacity to suit means to ends. It also must take into account the "Oedipus effect," that is, self-fulfilling prophecies, or looping effects, namely, altered courses of action, and hence outcomes, due to actors' changing knowledge and ideas.

Friedman and Kuhn's Alternatives

When Popper translated and published his *magnum opus* on philosophy of science in 1959, it was well received by philosophers of science and caused ripples in economics. Hayek had long said that he followed Popper; so had Hutchison and Robbins. Initial enthusiasm was followed by worried retreat. The enthusiasm was because Popper offered a clear criterion of scientific character, testability, so economists could legitimate their work as science by developing testable theories and testing them. Immediately, mathematical economists and others challenged this criterion on the grounds that their abstract models were a priori intellectual structures, not testable generalizations. Milton Friedman in 1953 had already offered a way out: Treat economic models instrumentally and judge them by their fruitfulness and

predictions. He ascribed to the neoclassical model of practical success. His was a philosophy of science variously called pragmatism and instrumentalism, which sacrificed truth as an aim and replaced it with *usefulness*. It fails as soon as we raise the question "Are claims of success true?" Popper, however, allowed that instrumentalism could be held without inconsistency, but he argued that the aim of science was to seek true explanations, not just useful knowledge. The latter was too narrow to capture the enterprise of scientists such as Galileo, Newton, and Einstein.

Not long after Popper's translation appeared, the American historian and philosopher of science Thomas S. Kuhn published his *The Structure of Scientific Revolutions*, in which he argued from historical materials that science differed significantly from the picture presented by Popper's philosophy. Kuhn took his inspiration from "Big Science," that is, science in large organized research units, often publicly funded and more or less modeled on the Manhattan Project for building the atomic bomb. Instead of valuing disputatiousness, such science valued consensus. In fact, if there was no consensus, there was no science, at least not efficient science as we know it. The process of achieving consensus was described by Kuhn in political terms. He described disputatiousness in politico-psychological terms as power-seeking persuasion. Adherents of a theoretical paradigm defended it as any political establishment will. But there comes a time when its anomalies create social instability and scientists cast around for an alternative. When consensus fails, research is at best pre- or proto-science. The next candidate for consensus emerges from the ranks of the leadership, or else Young Turks overthrow the old guard and establish the new paradigm themselves. Once consensus develops again, science resumes as researchers smooth out minor anomalies, try better to formulate the tenets of the new paradigm (especially by writing textbooks), and generate research programs that employ it.

Social scientists found Kuhn's model much more appealing than Popper's, since it had no normative dimension, or the normative dimension was submission to authority, as you will; it is naturalistic: Science is what scientists do. If social scientists differed from natural scientists in their lack of consensus, that was a difference, not a deficiency. (Indeed, Kuhn did not claim that his view applies to social science.) Popper and Kuhn (and their followers) had a number of sharp exchanges. Kuhn was equivocal about whether science pursued truth and made progress. Popper allowed that much of Kuhn's description of science was ethnographically accurate; this was because Big Science had become an uncritical, dogmatic, and rent-seeking enterprise that endangered the very existence of science as a self-critical quest for truth. Popper's model was Einstein and his theories of special and general relativity. Popper was a theorist of the best science being always spoiling for a new and earth-shaking revolution. Kuhn was a theorist of the bureaucratization of science, of its settling down into routine, what he famously dubbed "Normal Science." Kuhn thought puzzle-solving within an unquestioned paradigm was the state in which science performed best. Popper wondered where the big, world-shaking, exciting ideas would come from. Kuhn sought consensus; Popper relished dissensus. Kuhn's model was of conformity and productivity; Popper's was of rebellion and challenge. Yet the choice between them is not a matter of mere preference: Science is a central aspect of modern society, and its description is true or false. Kuhn did not deny Popper's characterization of some science as revolutionary but claimed that it was limited to periods of crisis, when it gets inefficiently mixed up with less respectable intellectual activities. He then admitted that there are big revolutions and small. This amounts to capitulation to Popper.

Ian Jarvie

See also Critical Rationalism; Deduction; Explanation, Theories of; Falsifiability; Induction and Confirmation; Kuhn and Social Science; Pessimistic Induction; Prophecy, Self-Fulfilling/Self-Defeating; Scientific Method; Situational Analysis

Further Readings

Friedman, M. (1953). The methodology of positive economics. In *Essays in positive economics* (pp. 3–46). Chicago, IL: University of Chicago Press.

Jarvie, I. (2002). *The republic of science: The emergence of Popper's social view of science, 1935–1945*. Amsterdam, Netherlands: Rodopi.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Popper, K. R. (1945). *The open society and its enemies*. London, England: George Routledge.

Popper, K. R. (1957). *The poverty of historicism*. London, England: Routledge. (Original work published 1943–1944)
Popper, K. R. (1959). *The logic of scientific discovery*. London, England: Hutchinson. (Original work published 1935 in German as *der Forschung*, Vienna, Austria: Springer)

Positivism, History of

Generally speaking, positivism is a label given to any philosophical position that embraces scientism, which is the view that all genuine knowledge is obtained by combining empirical observation and logical reasoning in a manner best exemplified by the natural sciences.

The idea of a "positive philosophy" was first formulated in the early 19th century by Auguste Comte, but its best-known version is Vienna Circle logical positivism, developed in Germany and Austria in the mid 20th century. The story of positivism is usually told backward, starting with the rise and fall of logical positivism, contrasting it with classical positivism, and concluding that our century is largely post-positivist. However, positivism is not entirely a phenomenon of the past; hence, one learns more by starting with what all positivists share.

All positivists agree that the most important fact about the modern world is the *success of science*. Only science obtains truly objective knowledge, independent of ideology and sociohistorical contingency. The reason for this success is the use, in every science, of some variant of a single method; hence, positivist philosophers must be epistemologists of science, articulating and defending this method. All positivists assume that the fundamental aim in acquiring objective knowledge is improvement of the human condition. Logical positivists deny that they espouse any philosophical doctrines and define their task entirely in epistemological terms, but this obscures the fact that their view, as much as their predecessors', is the intellectual expression of a broadly secular, antispeculative, science-promoting, politically reformist position. All positivists agree with Comte's slogan (emblazoned on Brazil's flag): Order and Progress: From knowledge of the natural and social order of things comes progress in successfully dealing with them.

Positivists also hold that this slogan remains true even when the "knowledge" embraced is prescientific. Thus, a world superstitiously understood leads to lives superstitiously lived. Comte offers an elaborate "Law of Three Stages" to explain this point. Under the right conditions, human intelligence goes through three developmental periods: (1) the theological (superstitious), (2) the metaphysical (ideological), and (3) the positive (scientific). This law characterizes human development from several angles: individual intellectual growth, world history, the rise of each science, and the emergence in human societies of religious, military, industrial, and legal activities suited to each stage.

Overall, Comte's law tells the story of our failure to explain reality in terms of the supernatural or natural powers hidden behind what we experience. Yet from these very failures, we learn to limit our search for knowledge to what we can actually observe. Every positivism is thus some species of *empiricism*. Reason should not be a slave to feelings or faith (as in theology); it cannot be its own authority (as in metaphysics). Mature minds abandon all efforts to solve life's ultimate mysteries (Why are we here? What happens after death?), confine their quest for knowledge to the study of observable phenomena, and develop a hierarchy of sciences—the last and most complex being *sociology* (a term Comte invented), which has special importance. It is certainly good to have power over material nature, but knowledge of ourselves as social beings facilitates an even higher purpose: establishing truly peaceful, flourishing societies.

Later positivists rejected Comte's law, in its details and as an empirical theory, but much of his "positivist spirit" passed through to them, even as they denied it. Émile Durkheim's early writings and John Stuart Mill's *Auguste Comte and Positivism* foreshadow the rigorous but still implicitly Comtean revisionism of logical positivism. They praise Comte for conceiving science methodologically, for promoting the idea of a distinctively "social" science and for extending the method of the natural sciences to it, and for appealing to the authoritative character of scientific knowledge to press for material and social reform. But they deem Comte's account of the rise of the scientific era irrelevant now that it has arrived and condemn his epistemology of science for its imprecision. Against the background of this critical appropriation, Comte's 20th-century progeny often rebranded themselves. Logical they surely are; but since they see history as philosophically irrelevant and base their knowledge

claims on science-minded observation strictly, not loosely, they are really logical "empiricists," not just positivists.

In its heyday, logical empiricism's hostility toward history (and the humanities and, of course, metaphysics), its formal-logical "rational reconstruction" of scientific reasoning, its narrow conception of "meaningful" utterances, and its picture of science as essentially cognitive and neutral constituted the gold standard for English-language philosophy—what was to become "analytic philosophy."

But by the 1970s, all its main features were being rejected by both scientists and philosophers. Yet if rejection of its explicit program is today virtually complete, the views that have replaced it are not entirely post-positivist. The legacy of classical positivism survives (a) in the widespread privileging of science and its technologies, (b) in the idea that human practices are at their best when they are known scientifically and follow sciencelike advice, (c) in the secular-progressivist construal of political economies, and (d) in the assumption that the whole drift of world history is necessarily toward the completion of what "we" call the "developed" world.

Robert Scharff

See also Empiricism; Explanation, Theories of; Logical Positivism/Logical Empiricism; Mill and the Moral Sciences; Philosophy of Sociology, History of; Scientific Method

Further Readings

Comte, A. (1830–1842). *Cours de philosophie positive* (6 vols.). Paris, France: Bachelier. (No complete English translation. See bibliographical details in "Comte, Auguste," *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/comte)

Comte, A. (1875–1899). *System of positive polity* (4 vols.; J. H. Bridges, Trans.). London, England: Longmans, Green. (Original work published 1851–1854)

Mill, J. S. (1969). Auguste Comte and positivism. In *Collected works* (Vol. 10). Toronto, Ontario, Canada: University of Toronto Press.

Neurath, O., Carnap, R., & Hahn, H. (1973). The scientific conception of the world: The Vienna Circle. In M. Neurath & R. S. Cohen (Eds.), *Empiricism and sociology* (pp. 299–310). Dordrecht, Netherlands: D. Reidel.

Pickering, M. (1995–2009). *Auguste Comte: An intellectual biography* (3 vols.). Cambridge, England: Cambridge University Press.

Richardson, A., & Uebel, T. (Eds.). (2009). *Cambridge companion to logical empiricism*. Cambridge, England: Cambridge University Press

Scharff, R. C. (1995). *Comte after positivism*. Cambridge, England: Cambridge University Press.

Simon, W. M. (1963). *European positivism in the nineteenth century: An essay in intellectual history*. Ithaca, NY: Cornell University Press.

Sorrell, T. (1991). *Scientism: Philosophy and the infatuation with science*. London, England: Routledge.

Postcolonial Studies

The entry introduces the blossoming, transdisciplinary, recent field of postcolonial studies, critically reviews its main elements and theoretical presuppositions, and shows its relevance to a renewed type of social science.

Postcolonial (often "post-colonial") studies deal with the effects of colonization on cultures and societies. Unlike *colonialism*, the term *postcolonialism* is of recent provenance, emerging in the late 1970s to describe a range of literary and cultural analyses of colonized and formerly colonized societies. As originally used by historians after World War II in terms such as the *postcolonial state*, "postcolonial" had a chronological meaning, designating the postindependence period. However, from the late 1970s, "postcolonial studies" has referred to the analysis by cultural critics of the various social, political, and cultural effects of colonization. The term *postcolonial* now describes neither a historical period nor a fixed range of societies but is best understood as a discourse generating a specific reading practice.

Origins

The study of the controlling power of representation in colonized societies had begun in the 1950s with the work of Frantz Fanon, and it reached a climax in the late 1970s with Edward Said's *Orientalism*. This study led to the development of colonialist discourse theory in the work of critics such as Gayatri Chakravorty Spivak and Homi Bhabha, although the actual term *postcolonial* was not employed in these early studies of the power of colonialist discourse to shape and form opinion and policy in the colonies and metropolises. While the analysis of the effects of colonial representation were central to

the work of these critics, the term *postcolonial* itself was first used to refer to cultural interactions within colonial societies in literary circles (see, e.g., the 1977 special issue of *New Literature Review*—the work of Ashcroft, Cotter, Docker, & Nandan). This was part of an attempt to politicize and focus on the concerns of fields such as Commonwealth literature and the study of the so-called New Literatures in English, which had been initiated in the late 1960s. *The Empire Writes Back* by Ashcroft, Griffiths, and Tiffin offered the first systematic account of the theoretical issues generated by postcolonial literatures and initiated the field that became known as *postcolonial literary studies*. The term has subsequently been widened to the field of *postcolonial studies*, which analyzes the political, linguistic, social, and cultural experience of societies that were former European colonies.

Postcolonialism and Globalization

The remarkable rise of postcolonial studies in the 1990s was partly due to the fact that by the late 1980s the world was hungry for a language to describe the diversity of cultures and the intersecting global range of cultural production. Postcolonial theory provided that language, a way of talking about the engagement of the global by the local, particularly local cultures, and, most importantly, provided a greatly nuanced view of globalization, developed from its understanding of the complexities of imperial relationships. According to critics such as Simon Gikandi, a substantial overlap existed between postcolonialism and globalization studies in that they shared at least two important elements: Both kinds of studies seek to explain forms of social and cultural organization that aim at transcending the boundaries of the nation-state, and they are also both concerned with offering new standpoints for understanding cultural flows that a homogeneous Eurocentric narrative of development and social change cannot explain.

It can be argued that the language of postcolonial studies drove a cultural turn in globalization studies in the 1990s, and it did this for three reasons. First, the systematization of postcolonial theory occurred at about the same time as the rise to prominence of globalization studies in the late 1980s. Second, it was around this time that literary and cultural theorists realized that debates on globalization had become bogged down in the classical narrative of modernity. Third, it became clear that there were many globalizations and that, far from the homogenizing downward pressure of economic globalization and the Washington consensus, a circulation of local alternatives could be seen to affect the nature of the global. It was through *cultural* practices that difference and hybridity, diffusion and the imaginary, concepts that undermined the Eurocentric narrative of modernity, were most evident.

Disciplinary Contests

The term *postcolonialism*, however, was a potential site of disciplinary and interpretative contestation almost from the beginning, especially the implications involved in the signifying hyphen or its absence. The heavily poststructuralist influence of the major exponents of colonial discourse theory—such as Edward Said (and Michel Foucault), Bhabha (and Louis Althusser and Jacques Lacan), and Spivak (and Jacques Derrida)—led many critics, concerned to focus on the material effects of the historical condition of colonialism, as well as on its discursive power, to insist on the hyphen to distinguish postcolonial studies *as a field* from colonial discourse theory per se, which formed only one aspect of the many approaches and interests that the term *postcolonial* sought to embrace and discuss. The hyphen has also been used as a reminder that postcolonialism is not a master discourse but a plethora of overlapping discourses concerned with the cultural effects of imperialism.

While this distinction has been largely superseded, the term *postcolonial studies* has come to refer to the wide and diverse analysis of European territorial conquests, the various institutions of European colonialisms, the discursive operations of empire, the subtleties of the construction of the subject in colonial discourse as well as the resistance of those subjects, and, most importantly perhaps, the differing responses to such incursions and their contemporary colonial legacies in both pre- and postindependence nations and communities. While postcolonial studies have tended to focus on the cultural production of such communities, they have expanded to include historical, political, sociological, and economic analyses, as these disciplines continue to engage with the impact of European imperialism upon world societies.

Historical Boundaries

The prefix *post* in the term also continues to be a source of vigorous debate among critics. The simpler

sense of *post* as meaning "after" colonialism has been contested by a more elaborate understanding of the working of postcolonial cultures that stresses the articulations between and across the politically defined historical periods of precolonial, colonial, and postindependence cultures. *The Empire Writes Back* used the term to refer to all the literature written after colonization. Thus, the term *postcolonial* is best understood as a reading practice that analyzes the continuing resistances, appropriations, and transformations of dominant ("imperial") discourses, institutions, and methodologies by colonized and formerly colonized societies. As a result, further questions have been asked about what limits, if any, should be set round the term. Aijaz Ahmad complains, for instance, that when the term *colonialism* can be pushed back to the Incas and forward to the Indonesian occupation of East Timor, then it becomes "a transhistorical thing," something that is always present but also always in a process of dissolution in some parts of the world. It is clear, however, that *postcolonial studies*, as the term has been employed in most recent accounts, have been primarily concerned to examine the processes and effects of, and reactions to, European colonialism from the 16th century up to and including the neocolonialism of the present day.

Locality and Materiality

An equally fundamental constraint is attention to precise location. Colonialism itself is much more complex and problematic than it seems at first: It is by definition transhistorical and unspecific, and it is used in relation to very different kinds of historical oppression and economic control. Every colonial encounter or "contact zone" is different, and each "postcolonial" occasion needs, against these general background principles, to be precisely located and analyzed for its specific interplay. A vigorous debate revolved around the potentially homogenizing effect of the term in the early 1990s. The effect of describing the colonial experience of a great range of cultures by this term, it was argued, is to elide the differences between them. However, there is no inherent or inevitable reason for this to occur. The materiality and locality of various kinds of postcolonial experience provide the richest potential for postcolonial studies, and they enable the specific analysis of the various effects of colonial discourse.

The theoretical issues latent in these two fundamentals—materiality and location—lie at the basis of much of the dispute over what the term *postcolonial* refers to and what it should or should not include. Yet, despite these disputes and differences, signs of a fruitful and complementary relationship between various postcolonial approaches have emerged in recent work in the field. Whether beginning from a basis in discourse theory or from a more materialist and historical reading, most recent discussions have stressed the need to retain and strengthen these fundamental parameters in defining the idea of the postcolonial. As critics like Robert Young have indicated, the crucial task has been to avoid assuming that the reality of the historical conditions of colonialism can be safely discarded in favor of 'the "fantasmatics of colonial discourse." On the other hand, as Young also warns, although the totalizing aspects of discourses of the postcolonial are of real concern, it is necessary to avoid a return to a simplified form of localized materialism that refuses entirely to recognize the existence of general discourses of colonialism and their effect on individual instances of colonial practice.

The challenge for many postcolonial critics and theorists lay in the need to strike a proper balance between the specific cultural and social conditions of colonized and formerly colonized peoples and the larger theoretical frameworks in which the cultural practices of those societies are analyzed. Not every colony will share every aspect of colonialism, nor will it necessarily share some essential feature since, like any category, it is—to use Ludwig Wittgenstein's metaphor—a rope with many overlapping strands. Despite the material differences of particular postcolonial peoples, each must deal with a common discursive, social, or political media provided or imposed by colonialism itself. Postcolonialism in general is an engagement, in all its many varieties, with these media. For instance, the ways in which postcolonial peoples have appropriated and transformed metropolitan languages may differ in particular instances, but the general political processes by which this appropriation occurs show the continuing value of broader postcolonial frameworks.

Diversity of Themes

A major feature of postcolonial studies has been their ability to analyze a vast array of cultural developments: race and racism, expressions of anticolonial nationalism, the paradoxical dissolution of the idea of nation along with the continuous persistence of national concerns, the question of language and

appropriation, the question of the transformation of literary genres, the question of ethnicity and its relation to the state, the increasing mobility of formerly colonized populations, and the impact of diasporas upon the social complexion of metropolitan centers. Since their emergence in the late 1980s, postcolonial studies have widened their reach to include issues such as the place of translation; the space of the sacred; the field of biopolitics, ecological theory, and animal rights; the role of the city; and, most importantly, the relation between local communities and global influences. This relationship is one for which postcolonial studies have long been prepared, both from their involvement in the cultural turn in globalization studies and from their long interest in the relation between local differences and the broader global impact of imperialism. An aspect of this outward movement is the rapidly growing interest in the mobility of formerly colonized populations, the movement of refugees and asylum seekers, and the characteristics of diasporas of various ethnic groups.

Thus, the future of the field seems set to move further out of the realm of classical imperialism and its cultural and political effects and into a growing investment in globalization studies, diaspora, transnationalism, and cosmopolitanism. The historical experience of imperial and colonial relations appears to provide valuable insights into the agency of local communities and the cultural production of subaltern, marginalized, and exiled peoples in a global era.

Bill David Ashcroft

See also Cultural Studies; Discourse Analysis; Enlightenment, Critique of; Foucault's Thought; Modernity; Norbert Elias: Process of Civilization and Theory of Sciences; Postmodernism; Race, Theories of

Further Readings

Ahmad, A. (1995). The politics of literary postcoloniality. *Race and Class, 36*(3), 1–20.

Appadurai, A. (1996). *Modernity at large: Cultural dimensions of globalization.* Minneapolis: University of Minnesota Press.

Ashcroft, B. (1996). On the hyphen in post-colonial. *New Literatures Review, 32,* 23–32.

Ashcroft, B., Griffiths, G., & Tiffin, H. (1989). *The empire writes back: Theory and practice in post-colonial literatures.* London, England: Routledge.

Ashcroft, W. D., Cotter, M., Docker, J., & Nandan, S. (Eds.). (1977). Post-colonial literature [Special issue]. *New Literature Review, 2.*

Chrisman, L. (2003). *Postcolonial contraventions: Cultural readings of race, imperialism, and transnationalism.* Manchester, England: Manchester University Press.

Desai, G., & Nair, S. (2005). *Postcolonialisms: An anthology of cultural theory and criticism.* New Brunswick, NJ: Rutgers University Press.

Gikandi, S. (2001). Globalization and the claims of postcoloniality. *South Atlantic Quarterly, 100*(3), 627–658.

Goldberg, D. T., & Quayson, A. (2002). *Relocating postcolonialism.* Malden, MA: Blackwell.

Hawley, J. C. (2001). *Encyclopedia of postcolonial studies.* Westport, CT: Greenwood.

Loomba, A. (Ed.). (2005). *Colonialism–postcolonialism.* New York, NY: Routledge.

Quayson, A. (2000). *Postcolonialism: Theory, practice or process?* Cambridge, England: Polity Press.

Spivak, G. C. (1999). *A critique of postcolonial reason: Toward a history of the vanishing present.* Cambridge, MA: Harvard University Press.

Venn, C. (2006). *The postcolonial challenge: Towards alternative worlds.* London, England: Sage.

Wisker, G. (2007). *Key concepts in postcolonial literature.* Basingstoke, England: Palgrave Macmillan.

Young, R. (2001). *Postcolonialism: An historical introduction.* Oxford, England: Blackwell.

Postindustrial Society

There are perhaps as many definitions of postindustrial society as there are of industrial society. Yet at the heart of all definitions has been the idea that the evolution of industrialism was in the process of undermining its traditional classic forms. In this sense, speculations on postindustrial society share many similarities with subsequent discussions and debates over *post-Fordism* and *postmodernism.* The concept and theory of postindustrial society inherited the classical language of sociology, forged in the cultural trauma surrounding the birth of modern industrial capitalism. It invoked the language of a Great Divide and portrayed sociology as the witness, harbinger, and prophet of a new, unborn age.

The continuing relevance of discussions of postindustrial society does not rest so much in its somewhat dated 1970s and 1980s analysis of the rise of a knowledge-based service society, the characteristics

of which are outlined in the first part of this entry. As outlined in the second section, it lies, rather, in its expression and embodiment of a particular grand narrative, portraying a particular view of present society as on the brink of a radically new era. The weaknesses of this discourse are addressed in the conclusion.

The Earlier Analysis

The idea of postindustrial society was popularized in academic circles by Daniel Bell in his book *The Coming of Post-Industrial Society*. Bell gave particular significance to the transition from a goods-producing economy to a service economy; the increasing predominance of white-collar and professional occupations; the development of electronics and the growth of automatic, computer, and communication technologies; the increasing role of theoretical or scientific knowledge; and widespread affluence. Bell argued that these developments were leading to the increasing economic centrality of *knowledge*, the university and white-collar professional occupations and ethics, the transformation of the corporation into a social institution, and a culture that assigned increasing priority to self-expression, equality, participation, and the general quality of life.

In contrast to Bell's thesis, radical critics of postindustrial society regarded such trends within a knowledge-based service society as an extension rather than a transcendence of key principles of industrialization such as centralization, bureaucratization, specialization, the pursuit of economic growth, and the continued development of science in the pursuit of more extensive control of nature and society. True postindustrialism, they argued, should be defined in opposite terms, through the emerging potential for decentralization, the demythologization of science, the recognition of limits to growth, growing public participation, and the challenge to traditional, scientific, and professional authority.

Despite their differences, such theories of postindustrialism came under fire for exaggerating the extent and significance of the trends they described. A particular area of contention was the significance of the role and power attributed to theoretical knowledge and those who possessed it, the homogeneous characteristics that were attributed to white-collar and professional work, and the effect this had on class consciousness, social values, and the political system.

Criticizing the Debate

The details of the debate are, however, less significant than the general terms in which it was carried out. Similar to previous debates over the rise of postcapitalist societies, a central weakness was the explicit or implicit technological and economic determinism embodied in such theories and their one-directional evolutionary approach to social change. The speculations over postcapitalist and then postindustrial society shared the same one-dimensional, apocalyptic, and universalistic approach to social transformation and the "issues of the age" that characterized the work of much of the writings of the 19th-century founding fathers of sociology.

From the 18th-century Enlightenment onward, intellectuals in Western society have grappled with the nature and implications of the development of reason and the establishment of progress in human affairs. Traditional speculations on the rise of industrial society, and subsequently postindustrial society, have tended to restrict the sociological imagination to a particular limited perspective on such issues. This has occurred in three main ways. First, industrial and postindustrial society theories have focused attention on the nature and implications of *scientific* reason and *technical and economic* progress. This contrasts and clashes with alternative critical and radical traditions that emphasize the development and significance of alternative forms of reason to that of science, the presence or removal of social domination, and the establishment of more democratic societies within alternative paths of development. Second, industrial and postindustrial society theorists tended to view the main transition in history as either having already occurred or being the result of ongoing evolution of existing trends. Little or no recognition was given to the possibility of, or requirements for, reason and progress taking the form of radical political transformation, of the kind advocated by revolutionary theorists. Third, social transformation was seen in universal and deterministic terms as the inevitable working out of a logic of reason or social development, possessing similar requirements and creating similar impacts throughout the world, with relatively less significance given to the possibility or value of alternative trajectories in different cultures, eras, or locations throughout the world.

Concluding Remarks

It is possible to observe the continuing influence of such restricted views of the sociological imagination in many contemporary grand narratives of late-modern societies, viewing such societies as undergoing a process of transition to post-Fordism or postmodernism. While current debates over postmodernity are more complex, sophisticated, and multifaceted than previous discussions of postindustrial society, the fate of speculation on postindustrial societies should warn us of the temptation and dangers of adopting a too one-dimensional, evolutionary, and deterministic view of social transformation and the central issues of the age.

Richard Badham

See also Enlightenment, Critique of; Experimenting Society, The; Frankfurt School and Critical Social Theory; Information Society; Knowledge Society; Postcolonial Studies; Postmodernism

Further Readings

Badham, R. (1986). *Theories of industrial society.* London, England: Croom Helm.

Bell, D. (1973). *The coming of post-industrial society: A venture in social forecasting.* New York, NY: Basic Books.

Kumar, K. (1979). *Prophecy and progress: The sociology of industrial and post-industrial societies.* Harmondsworth, England: Penguin.

Touraine, A. (1971). *The post-industrial society. Tomorrow's social history: Classes, conflicts and culture in the programmed society.* New York, NY: Random House.

Postmodernism

Postmodernism may be defined in the broadest sense as a crisis of representation. The concept seems to have originated in literature and the arts in the late 19th century, when it was used to describe various aesthetic movements and techniques that were seen by their advocates as somehow superseding the modernist orthodoxy.

Philosophy

In philosophy, postmodernism takes the form of a radically antifoundationalist approach that rejects all attempts to provide a universal grounding for knowledge and truth. Associated most closely with the French philosophers Gilles Deleuze, Michel Foucault, Jean-François Lyotard, and Jacques Derrida, postmodern antifoundationalism is based largely on two sources: (1) a controversial interpretation of the work of the 19th-century German philosopher Friedrich Nietzsche and (2) the poststructuralist critique of structural linguistics. For French postmodern philosophy, Nietzsche's radical, anti-Enlightenment critique of reason negates the possibility of universal scientific knowledge in any form, leading to a radical perspectivism where there are no such things as facts but only interpretations. Structural linguistics was arguably the dominant paradigm in French intellectual thought during the 1960s. Based on the original analysis by the Swiss linguist Ferdinand de Saussure, the goal of structural linguistics, which Saussure termed *semiology*, was to identify a universal set of linguistic rules that underlay all particular forms of speech. The poststructuralist critique, however, rejects the notion of an underlying structure in favor of an emphasis on the arbitrary nature of linguistic signs, also a concept taken from the work of Saussure.

Social Science

While the impact of postmodernism in American philosophy has been minimal, it has been far greater among the social sciences, especially sociology and anthropology, beginning in the 1980s with the translation into English of the works of the major postmodern French philosophers.

Four distinct usages of the concept of postmodernism may be identified: (1) as a substantive-historical category, (2) as a theoretical term, (3) as a methodological approach, and (4) as an ethico-political concept.

1. As a substantive category, postmodernism denotes a series of structural transformations in advanced industrial societies of sufficiently far-reaching nature and scope to suggest the rise, or at least transition toward, a new historical era: "postmodernity." Dated by some as beginning after World War I, by others after World War II, and by yet others since the 1960s, postmodernity is characterized by fundamental changes in virtually every sphere of contemporary life, including major institutions such as the economy, the polity, the family,

religion, science, architecture, and the arts, as well as those shaping core aspects of the self, including the nature of gender, racial, and sexual identities. Primary emphasis is placed on the role of new transportation and communication technologies, especially the airplane, the mass media, and the computer.

Most social scientists whose work is focused on postmodernism as a category describing contemporary social and cultural changes do not consider themselves postmodernists, however. Many of those who study the institutional changes that define postmodernity are critical theorists or Neo-Marxists who situate the development of new technologies within the broader context of a more advanced political-economic system, variously termed *global*, *digital*, or *multinational capitalism*. Similarly, many of those who characterize contemporary changes in the nature of self and identity by employing terms such as *the mutable*, *the saturated*, the *mass-mediated*, or *the fragmented self* base their analyses on conventional sociological and social-psychological approaches.

2. Others, however, embrace the notion, first suggested by the maverick sociologist C. Wright Mills, that new times call for different tools. Thus, as a theoretical term, *postmodernism* implies a renunciation of universal theories in favor of more particularistic approaches. In so doing, postmodern theories in the social sciences adopt the position of the French philosophers, especially that of Lyotard, who calls for a rejection of the grand "metanarratives" of universal science and political emancipation that characterized modern thought, in favor of more limited, local narratives. Postmodern theorists also employ the deconstructionist approach of Jacques Derrida to expose the constructed, arbitrary character of social categories such as gender, race, and sexual identity.

3. Postmodern research methods in the social sciences fall generally into two categories. The first apply poststructural strategies for the reading of literary texts, heavily influenced by the work of Foucault and Derrida, to a variety of social settings and, more controversially, to films and other mass media. These studies focus variously on revealing unintended, ideological, and/or multiple meanings in social settings and mass media texts. The second version of postmodern research methods employs the poststructuralist critique of texts and authorship to challenge the status of conventional ethnographic accounts in anthropology and sociology. Instead, postmodern ethnographers turn to models from literature and the arts in developing new or experimental ethnographic styles that, in their view, more authentically represent the groups they study. These include autoethnographies, dialogic or polyphonic ethnographies, ethnographic poetry, and ethnographic performances.

4. Finally, as a normative or ethico-political concept, postmodernism implies a renunciation of all claims to universal ethical and political principles in favor of an emphasis on pluralism and diversity. Criticizing the failure to recognize the rights of marginalized groups historically in the name of modern Enlightenment concepts of science and reason, postmodernists call for a new respect for difference without the need for transcendental grounding. To avoid falling into the opposite error of moral or political particularism, postmodernists advocate a strategy of coalition politics based on common interests with regard to specific issues.

Critical Remarks

Despite the incorporation of several aspects of postmodernism into some sectors of the contemporary social sciences, it remains a controversial and highly contested term. Criticisms of postmodernism in philosophy and the social sciences are largely organized along paradigmatic lines. For positivists, the postmodern rejection of the universally scientific project of the Enlightenment is intellectual heresy and is further seen as a dangerous form of irrationality. For interpretive social scientists, postmodernism is regarded as frivolous and narcissistic, especially in its advocacy of the new, experimental versions of ethnography. Finally, for critically oriented researchers, the postmodern deconstruction of social categories such as gender, race, and sexual orientation goes too far, undermining the basis for solidarity among marginalized groups that is seen as necessary for progressive social change.

David R. Dickens

See also Contemporary French Philosophy and the Social Sciences; Cultural Studies; Enlightenment, Critique of; Foucault's Thought; Modernity; Multiculturalism;

Postcolonial Studies; Postindustrial Society; Structuralism and Poststructuralism

Further Readings

Best, S., & Kellner, D. (1991). *Postmodern theory: Critical interrogations*. New York, NY: Guilford Press.
Dickens, D. R., & Fontana, A. (Eds.). (1994). *Postmodernism and social inquiry*. New York, NY: Guilford Press.
Jameson, F. (1991). *Postmodernism, or, the cultural logic of late capitalism*. Durham, NC: Duke University Press.
Lyotard, J.-F. (1994). *The postmodern condition: A report on knowledge*. Manchester, England: Manchester University Press.
Malpas, S. (2005). *The postmodern*. London, England: Routledge.
Marcus, G., & Fischer, M. (1986). *Anthropology as cultural critique: An experimental moment in the human sciences*. Chicago, IL: University of Chicago Press.
Rosenau, P. (1992). *Post-modernism and the social sciences*. Princeton, NJ: Princeton University Press.

Power

Power in its most generic sense simply means the capacity to bring about significant effects: to effect changes or prevent them. The effects of social and political power will be those that are of significance to people's lives. When these effects of power are such as to affect people's interests adversely, we speak of power being held or exercised *over* them—and the social scientist's quest is to try to reveal what this involves. There are other ways of identifying social and political power: for instance, as *collective* power to achieve shared goals (as when people cooperate to promote a cause or pursue a campaign) or as *positive* or beneficent power, where power serves others' interests (as, ideally, parents, teachers, philanthropists, and social workers are supposed to do).

This entry focuses on power exerted over others—while noting that power over others will often involve collective power and be combined with beneficent power. Power, thus understood, is arguably the more effective the less perceptible its workings are to agents and observers alike, thereby posing a problem for its social-scientific study. Four conceptions of power are outlined, each claimed by its proponents to reveal more of the phenomenon than its predecessor.

The One-Dimensional View

Power is most visibly at work when in a conflict of interests between agents (individual or collective) one prevails over another or others. By *exercising* power, the powerful agent demonstrably *has* that power, but since power is a capacity, that agent can also have the power without exercising it. The conflict, on this view, is between agents' overt preferences as revealed by their chosen behavior and can be interpersonal, within or between organizations, or between countries. It can be legitimate or illegitimate, or legitimacy may be what is at issue. And the winning agent may prevail through the rules of the game (as in political decision making or economic markets) or by threats or the offer of rewards.

Thus, the political scientist Robert Dahl and his colleagues sought to test the thesis of a ruling-power elite by investigating whether in New Haven, Connecticut, the preferences of the hypothetical ruling elite regularly prevail. They concluded that there was no power elite, since power, in this first sense, was distributed *pluralistically*, with different groups prevailing over different key issues.

The Two-Dimensional View

Critics—such as Peter Bachrach and Morton Baratz—objected that this conclusion resulted from seeing power as only manifested in behavior that concerns key issues over which there is observable, overt conflict, where that conflict is between the interests of the parties and where their interests consist in their conflicting observable preferences. Power, they claimed, is also at work when the powerful can, with more or less deliberate intent, suppress or thwart challenges to their interests by agenda control—that is, by deciding what gets decided, for example, by denying the *grievances* of marginal or excluded groups a hearing, through a variety of means, from censorship to the manipulation of procedures, thereby preventing potential demands from becoming actual ones.

This view of power relaxes the requirement that the conflict between parties' interests be manifested in overt behavior: Power must still involve conflict, but the conflict can be covert, involving that between dominant interests and the grievances of

the excluded or marginalized that have thus failed to be publicly heard. In the face of acquiescence, with no observable conflict between parties, there will, on this view, be no way to determine empirically whether their interests are opposed and thus whether consensus is genuine or has been attained through power.

The Three-Dimensional View

This issue has led to a further view, which, it has been claimed by Steven Lukes, incorporates the previous two but enables one to see further and deeper into the phenomenon. According to this view, power can indeed be at work in ways that are hidden from the view of those subject to it and even of its possessors. The powerful may work to avert conflict by contributing (intentionally or unintentionally) to getting others to want what they want them to want, shaping their perceptions, cognitions, and thus preferences. Of course, those subject to such power are typically themselves active participants or subjects, not just its objects or victims. Here, power is not just the ability to prevail over others in conflicts of interests and set the agenda of what such conflicts are about, it also encompasses being able to secure their dependence, deference, allegiance, or compliance, even without needing to act and in the absence of conflict.

This view rejects three assumptions implicit in the foregoing views. The first is that power must involve positive, intended actions by the powerful. Those subject to their power may anticipate their reactions and so engage, for instance, in self-censorship. They may defer or be attracted to the powerful and regard their subordinate place in the social order as self-evident, natural, unchangeable, and even beneficial. Pierre Bourdieu calls such power "symbolic power" and unreflective dispositions, and so to regard it as "habitus." A second assumption is that since power shows up where there is conflict, conflict is essential to power relations. But sometimes the most effective form of power is the capacity to prevent conflict from arising in the first place. The third (most problematic and contested) assumption is that this form of power can deflect people from understanding and thus pursuing their own interests. It can induce them to accept mythical and simplistic beliefs, playing on their fears, prejudices, and limited information and their disposition to engage in faulty reasoning. So the power to frame issues and thus help shape beliefs can be the power to mislead, misleading people to support leaders and follow policies that work against their interests.

Difficulties and Alternatives

But how are we to know, in the absence of observable conflict, whether or not the apparent acceptance and compliance that we observe has been secured by informational and cognitive power—the power to frame issues and thus help shape people's beliefs? More particularly, how would we ascertain whether the consensus results from what we have called *the power to mislead*, namely, misleading people to support leaders and favor policies that work against their interests?

There are various approaches within the social sciences to these questions. One is to avoid them. This is the approach of mainstream neoclassical economics and of those areas of political science influenced by it and by rational choice theory. This approach takes individuals' preferences as given, ignoring altogether the question of how they are formed. Within sociology, a roughly similar approach is adopted by those who take "value consensus," when they find it, as given, without raising the question of the extent to which it may in turn be shaped by power. That question is most directly raised within the Marxist tradition, in the form of what has been called the *dominant-ideology thesis*. According to this, as Karl Marx wrote in 1845, the ideas in every epoch are the ideas of the ruling class: The thesis was developed most explicitly by the 20th-century Italian Marxist Antonio Gramsci, who employed the notion of *hegemony* to refer to the inculcation of such ideas across civil society in subordinate classes that affirm, believe, and follow them in "normal times," when their conduct is not independent and autonomous but submissive and subordinate.

That thesis has been challenged, for example, by James Scott, who, focusing on studies of slavery, serfdom, untouchability, racial domination, including colonialism and highly stratified peasant societies, and also of total institutions, such as jails and prisoner-of-war camps, maintains that the victims of domination are in a state of constant rebellion. They are tactical and strategic actors who dissemble in order to survive. The appearance of consent and unanimity is for the public stage. Behind the scenes,

they voice their dissent in religious and ritual life. The subordinate group is constantly devising strategies to thwart and reverse the attempts of the powerful to maintain and exert their material control and symbolic reach.

An alternative account of quiescence, expounded by Jon Elster, is in terms of *adapative preferences*, where agents themselves display a disposition to accommodate their desires to what they view as feasible. This psychological mechanism can, of course, be adduced and encouraged by power, as when the underdog learns throughout a lifetime of socialization to endure the burdens of life by conformism and cheerful endurance, as in the case of traditional Indian women, as described by Martha Nussbaum. Profoundly dependent on males, often subject to domestic violence and denied opportunities for promotion and the learning of skills available to men, they can lack any sense of being wronged and regard their situation as natural and normal.

A Fourth View: Michel Foucault

The three previous views all agree in seeing power of some over others as power relations, and they assume that such power relations link *subjects*, or agents—that is, individuals or collective agents, such as groups or institutions—that are capable of choice, reflection, and the pursuit of strategies to achieve their goals. But not all students of power share this assumption. One such was Michel Foucault, whose view of power—often expressed in a rhetorical and exaggerated way—has been hugely influential across many disciplines.

Foucault agreed that power is at its most effective when least observable, but his focus shifted away from individuals and groups that dominate and are dominated. He saw power as operating "through" individuals rather than against them, as "constituting" the individual, who is at the same time its vehicle. He wrote of power as producing "subjects," forging their character and "normalizing" them, rendering them capable of and willing to adhere to norms of sanity, health, sexuality, and other forms of prevailing propriety. He claimed that these norms are maintained by policing the boundaries between the normal and the abnormal and by continuous surveillance and self-surveillance. This was the central theme of his earlier work on discipline and punishment and its suggestion that we live in a "carceral" or "disciplinary" society. That work is famous for deploying the image of Jeremy Bentham's late-18th-century conception of an ideal prison, the Panopticon, where continuous surveillance (or even just the belief that it is occurring) suffices to induce in the inmates a state of permanent and conscious visibility that ensures the automatic functioning of power. His later work concerned what he termed *governmentality*, referring to the ways in which in modern societies various authorities administer populations, in which individuals shape their own selves, and in which these processes get aligned. Now, Foucault wrote of *bio-power*, meaning a "life-administering power" concerned with social science and statistics to "normalize," control, and regulate the life and health of populations.

Steven Lukes

See also Decision Theory; Foucault's Thought; Governmentality and Regime; Habitus; Ideology; Political Psychology

Further Readings

Bachrach, P., & Baratz, M. S. (1962). The two faces of power. *American Political Science Review, 56,* 941–952. (Reprinted in *Power: Critical concepts*, 3 vols., by J. Scott, Ed., 1994, London, England: Routledge)

Bourdieu, P. (1991). *Language and symbolic power: The economy of linguistic exchanges* (J. B. Thompson, Ed.). Cambridge, England: Polity Press.

Dahl, R. A. (1961). *Who governs? Democracy and power in an American city.* New Haven, CT: Yale University Press.

Dowding, K. (2011). *Encyclopedia of power.* Thousand Oaks, CA: Sage.

Elster, J. (1985). *Sour grapes: Studies in the subversion of rationality.* Cambridge, England: Cambridge University Press.

Foucault, M. (2000). *Power* (The Essential Works of Foucault, 1954–1984, Vol. 3; J. D. Faubion, Ed., R. Hurley, Trans.). New York, NY: New Press.

Gramsci, A. (1971). *Selections from the prison notebooks of Antonio Gramsci* (Q. Hoare & G. Nowell-Smith, Eds.). London, England: Lawrence & Wishart. (Original work published 1926–1937)

Haugaard, M., & Clegg, S. R. (2009). *The SAGE handbook of power.* London, England: Sage.

Lukes, S. (2004). *Power: A radical view.* Houndmills, England: Palgrave Macmillan.

Nussbaum, M. C., & Glover, J. (1995). *Women, culture and development*. Oxford, England: Clarendon Press.
Scott, J. (Ed.). (1994). *Power: Critical concepts* (3 vols.). London, England: Routledge.
Scott, J. C. (1990). *Domination and the arts of resistance: Hidden transcripts*. New Haven, CT: Yale University Press.
Wartenberg, T. E. (1990). *The forms of power: From domination to transformation*. Philadelphia, PA: Temple University Press.

Pragmatism

Pragmatism as a philosophical doctrine traces back to the academic skeptics in classical antiquity. Denying the possibility of achieving authentic knowledge (*epistêmê*) regarding the real truth, they taught that we must make do with *plausible information* (*to pithanon*) adequate to the needs of practice. This entry gives an overview of philosophical pragmatism, its development and reception, as well as its significance as a philosophical school with repercussions for scientific method.

Charles Sanders Peirce

Pragmatism as a developed philosophical position stems from the work of Charles Sanders Peirce. For him, pragmatism was primarily a theory of meaning, with the meaning of any concept that has application in the real world inhering in the relations that link experiential conditions of application with observable results. But by the "practical consequences" of the acceptance of an idea or a contention, Peirce meant the consequences for *experimental* practice—"experimental effects" or "observational results"—so that for him the meaning of a proposition is determined by the essentially positivist criterion of its experiential consequences in strictly *observational* terms. And, moving beyond this, Peirce also taught that pragmatic effectiveness constitutes a quality control monitor of human cognition—though here again the practice issue is that of *scientific* praxis and the standard of efficacy pivoting on the issue of specifically *predictive* success. Peirce developed his pragmatism in opposition to idealism, seeing that the test of applicative success can lead mere theorizing to stub its toe on the hard rock of reality. But his successors softened up the doctrine, to the extent that with some present-day "pragmatists" the efficacy of ideas consists in their mere *adoption* by the community rather than—as with Peirce—in the success that the community may (or may not!) actually encounter as it puts those ideas into practice.

William James

Although Peirce developed pragmatism into a substantial philosophical theory, it was William James who put it on the intellectual map in his enormously influential *Pragmatism: A New Name for Some Old Ways of Thinking*, published in 1907. However, James changed (and—as Peirce himself saw it—ruined) Peircean pragmatism. For where Peirce saw in pragmatism a road to impersonal and objective standards, James gave it a personalized and subjectivized twist. With James, it was the personal (and potentially idiosyncratic) idea of efficacy and success held by particular people that provided the pragmatic crux, not an abstracted community of ideally rational agents. For him, pragmatic efficacy and applicative success did not relate to an impersonalized community of scientists but to a diversified plurality of flesh-and-blood individuals. For James, truth is accordingly what reality impels and compels human individuals to believe; it is a matter of "what pays by way of belief" in the course of human activity within the circumbient environment, and its acquisition is an invention rather than a revelation. With James, the tenability of a thesis is determined in terms of its experiential consequences in a far wider than merely *observational* sense—a sense that embraces the affective sector as well.

Variations

One overarching and ironic fact pervades the divergent development of pragmatism, namely, that the doctrine can be seen either as a validation of objectively cogent standards or as a subverter of them. There is a harder, *objectivistic* pragmatism of the right, a Peircean pragmatism of "what works *impersonally*"—though proving efficient and effective for the realization of some appropriate purpose in an altogether person-indifferent way ("successful prediction," "control over nature," "efficacy in need fulfillment"). And there is a softer, *subjectivistic* pragmatism, a Jamesean pragmatism of "what

works for *X*"—in proving efficient and effective for the realization of a particular person's (or group's) wishes and desires. The objective pragmatists stand in the tradition of Peirce and include the philosophers F. P. Ramsey, C. I. Lewis, and Rudolf Carnap; the subjective pragmatists stand in the tradition of William James and include F. C. S. Schiller and, more recently, Richard Rorty. (John Dewey straddles the fence by going to a social interpersonalism that stops short of impersonalism.) Looking at James, Peirce saw subjective pragmatism as a corruption and degradation of the pragmatic enterprise since its approach is not a venture in validating objective standards but in *deconstructing* them to dissolve standards as such into the variegated vagaries of idiosyncratic positions and individual inclinations. And this is how objectivistic pragmatists view the matter down to the present day.

Meaning and Truth

In the hands of this founding father, pragmatism had two principal components: one as regards *meaning* and the other as regards *truth*. Peirce's *meaning pragmatism* encompasses a pragmatic view of the meaning of concepts and ideas. The crux, so Peirce (1931–1958) maintained, lay in the "pragmatic maxim": "To ascertain the meaning of an intellectual conception one should consider what practical consequences might conceivably result from the truth of that conception; and the sum of these consequences will constitute the entire meaning of the conception" (p. 9). Meaning, in sum, is as meaning does. As Peirce put it in his classic 1878 essay "How to Make Our Ideas Clear," "There is no distinction of meaning so fine as to consist in anything but a possible difference of practice" (p. 257). Peirce insisted that the prime function of our beliefs regarding the world is to commit us to rules for action—to furnish guidance for our behavior in point of what to think, say, and do—and, above all, to canalize our expectations in matters of observation and experiment in scientific contexts.

And much the same sort of story here told with respect to meaning holds also with respect to *truth*. Those theses are true whose implementation in *practice* "work out" by way of yielding success in matters of prediction and application. For Peirce, the best route to this distinction is the scientific method, whose rivals—evidence-ignoring tenacity, pious adherence to authority, a priori speculation, and the like—simply cannot compare with it in point of producing trustworthy results. True factual beliefs, ipso facto, are those that achieve *efficacy* by guiding our expectations, beliefs, and actions in satisfying ways—where specifically cognitive satisfactions are at issue. And they must achieve this on a systemic basis.

As Peirce (1931–1958) saw it, truth in *scientific* matters consists in those contentions that are "fated to be ultimately agreed to by all who investigate [scientifically]" (p. 407). It is what the community of rational inquirers is destined to arrive at in the end—the ultimate consensus of informed opinion among investigators committed to the principles of science. To his mind, it is not mere inquiring as such but *properly conducted* inquiring that must eventually get at the truth of things. As Pierce saw it, it is the scientific method—not the scientific doctrine of the day—that is crucial for rational inquiry. And he rejected an ideology of the "look to science for all the answers" sort for the same reason that he rejected dogmatism of any sort, because it is itself ultimately unscientific.

Themes and Developments

In a community of rational agents, there is bound to be a parallelism between applicative efficacy and substantiative justification. This circumstance has far-reaching ramifications, since pragmatism here becomes conjoined to *evolutionism*. And control is a pivotal factor here. To be sure, if a bounteous nature satisfied our every whim spontaneously, without effort and striving on our part, the situation would be very different. For then, the beliefs that guide and canalize our activities would generally not come into play—they would remain inoperative on the sidelines, never being "put to the test." There would then be no need for active (and thought guided) intervention in "the natural course of things" within an uncooperative (at best indifferent, at worst hostile) environment. But *as things stand*, we are constantly called upon to establish varying degrees of "control over nature" to satisfy even our most basic needs (to say nothing of our virtually limitless wants).

The *developmental* perspective and the pragmatic approach thus join together into a seamless whole. A continuous thread links together the entire tradition of realistic pragmatism in its conviction

that the ongoing work of an enduring community of rationally competent inquirers will be self-monitoring—that mistakes will be detected and reduced in the course of time. The guiding conviction is that the community will, over the course of time, learn how to improve its procedures of inquiry through the processes of inquiry itself so that rational inquiry is in this sense self-monitoring and self-corrective. And here, we cannot reasonably look on nature as a friendly collaborator in our human efforts, systematically shielding us against the consequences of our follies and crowning our cognitive endeavors with a wholly undeserved success that ensues for reasons wholly independent of any actual adequacy vis-à-vis the intended range of purpose. And this leads us back to Peirce's grounding insight. Our cognitive methods are able to earn credit as giving a trustworthy picture of the world precisely because they evolve under the casual pressure of that world. In sum, the methods of scientific inquiry stand superior in its claims to provide an appropriate inquiry method on grounds that are essentially pragmatic. And this pragmatic superiority of science as a cognitive venture in matters of effective description, explanation, prediction, and control both manifests and serves to explain its emergence in cognitive evolution by rational selection via considerations of pragmatic efficacy.

The development of modern pragmatism has seen a steady diffusion of John Stuart Mill's idea that practice is the crux of normativity, as, for example, his idea that desirability is simply a matter of being desired. This tendency was clearly manifested in the thought of William James, for whom the meaning of language is to be equated with its manner of application, and the truth of claims, with their successful implementability in personal practice. And it is in much this spirit that John Dewey launched his assault on distinctions, in particular those between fact and value, between factual usability and normative appropriateness. This tendency continued in spades in the work of contemporary American pragmatists, as manifested, for example, in Richard Rorty's insistence on abandoning the whole armamentaria of concepts in classical philosophy. And we find it reemphasized in Hilary Putnam's identification of truth as nothing but the result of "our best practice" in inquiry.

Against this tendency, a more authentically Peircean realistic pragmatism would insist on construing objective normativity as something distinct from practical efficacy. Instead of taking the relation here as one of *identification*, it would be seen as merely a matter of *evidentiation*. Accordingly, one would not join Mill in seeing the condition of being desired as *constituting* desirability but rather merely (and vastly more plausibly) as *betokening* it. And we need not see truth as *constituted* by our inquiries but as merely *evidentiated* by them. In sum, the nondeconstructive pragmatism espoused here is not revisionary and reconstructive but merely *evidential*. And now the successful supplementation of our beliefs in application is merely the best *evidence* we can secure for the its implicative appropriateness as true, but it is not "the real meaning" of truth. Such an evidentialistic pragmatism does not see those dualistic distinctions, anathematized by Dewey, as conceptual illusions but rather as real gaps that can be crossed by the standard resources of evidential substantiation.

Regrettably, this is not how this line developed. As the 20th century progressed, pragmatism increasingly became a pathway to subjectivism.

Epistemic pragmatists want to reduce truth to warranted accessibility. Semantical pragmatists want to reduce meaning to conventionalized usage. Philosophical pragmatists want to substitute subjective intentions for objective norms. Historist pragmatists want to replace objective fact with what people think at the time. Metaphysical pragmatism wants to replace philosophical deliberation with interesting conversation. Where originally in Peirce's hands pragmatism provided an impersonally objective standard of judgment—to wit, purposive efficacy in application—now it became a way of loosening up to a position of subjectivistic and personalistic orientation.

Reception

Pragmatism has had a mixed reception in Europe. In Italy, Giovanni Papini and Giovanni Vailati espoused the doctrine and turned it into a party platform for Italian philosophers of science. In Britain, F. C. S. Schiller was an enthusiastic follower of William James, while F. P. Ramsey and A. J. Ayer endorsed pivotal aspects of Peirce's thought. Among Continental participants who flourished in the United States, Rudolf Carnap also put pragmatic ideas to work on issues of logic and

philosophy of language, and Hans Reichenbach reinforced Peirce's statistical and probabilistic approach to the methodology and prolification of induction. However, the reception of pragmatism by other philosophers was by no means universally favorable. The British idealist F. H. Bradley objected to the subordination of cognition to practice because of what he saw as the inherent incompleteness of all merely practical interests. The British philosopher G. E. Moore (an opponent of Bradley) also criticized William James's identification of true beliefs with useful ones—among other reasons because utility is changeable over time. Bertrand Russell objected that beliefs can be useful but yet plainly false. And various Continental philosophers have disapprovingly seen in pragmatism's concern for practical efficacy—"for success" and "paying off"—the expression of characteristically American social attitudes: crass materialism and naive populism. Pragmatism was thus looked down upon as a quintessentially American philosophy—a philosophical expression of the American go-getter spirit with its materialistic ideology of worldly success.

Nicholas Rescher

See also Abduction and Inference to the Best Explanation; Empiricism; Epistemology; Evolutionary Psychology; Idealism; Objectivity; Pragmatism and the Social Sciences; Truth, Philosophical Theories of

Further Readings

Brent, J. (1998). *Charles Sanders Peirce: A life* (Rev. ed.). Bloomington: Indiana University Press.

McDermid, D. (2006). *The varieties of pragmatism: Truth, realism, and knowledge from James to Rorty.* London, England: Continuum Press.

Murphy, J. P. (1990). *Pragmatism from Peirce to Davidson.* Boulder, CO: Westview Press.

Myers, G. E. (1986). *William James: His life and thought.* New Haven, CT: Yale University Press.

Peirce, C. S. (1878). How to make our ideas clear. *Popular Science Monthly, January,* 286–302.

Peirce, C. S. (1931–1958). Pragmatism. In C. Harsshorne & P. Weiss (Eds.), *Collected papers of Charles Sanders Peirce* (Vol. 5, pp. 9, 257, 407). Cambridge, MA: Harvard University Press. (Original work published ca. 1905)

Putnam, H. (1994). *Pragmatism*. Oxford, England: Blackwell.

Rescher, N. (1977). *Methodological pragmatism.* Oxford, England: Blackwell.

Rescher, N. (2000). *Realistic pragmatism.* Albany: SUNY Press.

Richardson, R. D. (2006). *William James: In the maelstrom of American modernism.* Boston, MA: Houghton Mifflin.

Rorty, R. (1982). *Consequences of pragmatism.* Minneapolis: University of Minnesota Press.

Shook, J., & Margolis, J. (Eds.). (2006). *A companion to pragmatism.* Oxford, England: Blackwell.

Smith, J. E. (1978). *Purpose and thought: The meaning of pragmatism.* New Haven, CT: Yale University Press.

Pragmatism and the Social Sciences

Originating in the United States in the 19th century, pragmatist philosophy has had a profound impact on the humanities and social sciences both within America and further afield. The first generation counted Charles Peirce, William James, and John Dewey among its ranks; more recent proponents have included Richard Rorty and Richard Bernstein. There are different forms of pragmatism, but most self-proclaimed pragmatists share an antifoundationalist stance in that they question the possibility of establishing temporal, universal criteria to judge ethical, aesthetic, or knowledge claims. Most pragmatists also oppose what Dewey called the "spectator theory of knowledge"—that is, the assumption that knowledge somehow passively represents the inner essence of the outer world. Instead, they see knowledge as an active intervention in the world. Finally, most pragmatists are reluctant to enter theoretical debates without a visible practical payoff.

Pragmatism initially developed at Harvard, but from the 1890s onward, the philosophy department at the newly founded University of Chicago became a vibrant center of pragmatist thought, eventually inspiring neighboring departments such as sociology. With key empirical works on various aspects of city life, the Chicago school of sociology became a hallmark for high-quality qualitative research on socially relevant topics. Among the earlier pragmatists, James and Dewey had a considerable influence on psychology and educational science, while George Herbert Mead had probably an even more significant impact on sociology. Educated at Harvard and in Germany, Mead taught in the

philosophy department at the University of Chicago, but he was not a prolific writer and only became influential in sociology posthumously through the interventions of former students Charles Morris and Herbert Blumer. Morris edited *Mind, Self, and Society* in 1934, a collection based on student notes of Mead's lectures on selfhood and social interaction, which has since become one of the canonical texts in sociology. Blumer was equally influential because he championed the sociological school of "symbolic interactionism," which was inspired by his former mentor's lectures on social psychology.

Mead's ideas became important for those sociologists who were interested in how people attribute meanings to their surroundings and act accordingly. Mead emphasized the distinctive nature of human interaction, which relied on symbols such as language. Because the meanings of symbols are shared, they enable participants in an interaction to adopt the perspective of the other participants. Thereby, the participants are able to predict what meaning these others would attribute to their possible "gestures" and to act accordingly. While Mead recognizes the power of the community (what he calls the "generalized other") over the individual, he insists that people's ability to adopt the attitude of others ties in with their capacity to transcend social, or indeed any, determination and thus to exercise genuine agency. In sociology, this account provided the theoretical building blocks for a research program that investigated the microfoundations of social life—that is, the interactions between individuals. Blumer's symbolic interactionism had particular affinities with Erving Goffman's dramaturgical approach and to a lesser extent with Harold Garfinkel's ethnomethodology. Albeit distinctive, all three approaches presented an alternative to the dominant structural-functionalist outlook at the time.

From the late 1940s and until the late 1970s, American philosophers increasingly focused on analytical philosophy at the expense of American pragmatism, but it is also during this period that Continental European philosophers, notably Jürgen Habermas, started to engage with pragmatism. While the first generation of the Frankfurt school was antagonistic toward pragmatism, Habermas turned toward Peirce to develop his critique of positivism. It is in this context that Habermas's 1971 work *Knowledge and Human Interests* emphasizes the interests that underscore distinct types of knowledge, with *empirical-analytical* knowledge directed toward control and prediction, *hermeneutics* tied in with understanding, and *critical theory* geared toward emancipation. Habermas's later *Theory of Communicative Action* (1984, 1987) also relied to a considerable extent on Mead's social theory. Mead's work, then, appeared as the theoretical basis for conceptualizing the potential of an open, unconstrained debate among equals.

With the publication in 1979 of *Philosophy and the Mirror of Nature*, the American philosopher Richard Rorty contributed to the revival of pragmatism in American philosophy, although it is particularly his later *Consequences of Pragmatism*, published in 1982, that elaborated on his indebtedness to Dewey. For Rorty, recent contributions to analytical philosophy have helped undermine the analytical enterprise as such, leading him to embrace the dialogical promise of hermeneutics instead of epistemology and its misguided promise of a neutral algorithm. Rorty only touches upon the social sciences, but some social theorists, such as Patrick Baert, have tried to demonstrate the relevance of this neopragmatist perspective for social research, arguing for self-referential research, whereby the encounter with difference potentially leads researchers to reevaluate the presuppositions that originally shaped their research. Several contemporary French social scientists share similar pragmatist concerns but draw inspiration from Dewey rather than Rorty. This is the case with Luc Boltanksi too, a former student of the French sociologist Pierre Bourdieu, who relies on Dewey to develop a sociology that maintains a notion of critique while recognizing the reflexivity and resourcefulness of people in everyday settings.

Patrick Baert

See also Ethnomethodology; Pragmatism; Self and the Social Sciences; Symbolic Interactionism

Further Readings

Baert, P. (2005). *Philosophy of the social sciences: Towards pragmatism*. Cambridge, England: Polity Press.

Blumer, H. (1969). *Symbolic interactionism: Perspective and method*. Englewood Cliffs, NJ: Prentice Hall.

Boltanksi, L. (2011). *On critique: A sociology of emancipation*. Cambridge, England: Polity Press.

Habermas, J. (1971). *Knowledge and human interests*. Boston, MA: Beacon Press.

Habermas, J. (1984). *The theory of communicative action: Vol. 1. Reason and the rationalization of society.* Boston, MA: Beacon Press.

Habermas, J. (1987). *The theory of communicative action: Vol. 2. Lifeworld and system: A critique of functionalist reason.* Boston, MA: Beacon Press.

Mead, G. H. (1934). *Mind, self, and society.* Chicago, IL: University of Chicago Press.

Rorty, R. (1979). *Philosophy and the mirror of nature.* Princeton, NJ: Princeton University Press.

Rorty, R. (1982). *Consequences of pragmatism* (Essays 1972–1980). Minneapolis: University of Minnesota Press.

PREFERENCE

The notion of preference is central to many social science disciplines. It also raises a number of philosophical issues central to the philosophy of social sciences. However, there is considerable disagreement about how to interpret this concept, with various incompatible accounts in use. This entry starts by sketching the main formal properties of the preference notion and their justification. It then addresses the relation of preference to choice, and the various interpretations offered.

The preference concept is often characterized as a binary relation defined over a set X of alternatives. In particular, there are two basic relations, strict preference ($\succ$) and indifference ($\sim$), and one derived relation, "at least as good as" ($\succcurlyeq$), defined as follows:

For all $A,B \in \boldsymbol{X}$: $A \succcurlyeq B$ if and only if $A \succ B$ or $A \sim B$ (weak preference).

In most applications, preferences are assumed to be *complete*. That is,

For all $A,B \in \boldsymbol{X}$: $A \succcurlyeq B$ or $B \succcurlyeq A$.

Completeness is sometimes difficult to justify. Indeed, we often do not have, and do not need, complete preferences. Consider a person who has to choose between three alternatives A, B, and C. If she knows that she prefers A to the others, she does not have to make up her mind about the relative ranking between B and C.

Incompleteness may arise for different reasons. Agents may not have a preference between two alternatives because they have not sufficiently thought about the alternatives (or lack sufficient information to do so). Thinking more about the issue will thus resolve incompleteness. Yet not all cases of incompleteness are resolvable. Even after maximum information intake and reflection, an agent may not be able to rank two alternatives. Such unresolvable incompleteness problems often arise when people consider alternatives to be *incommensurable*.

The other main assumption concerns the transitivity of preferences, which says that

For all $A,B \in \boldsymbol{X}$: If $A \succcurlyeq B$ and $B \succcurlyeq C$, then $A \succcurlyeq C$.

People often violate transitivity. The controversial question is whether transitivity should be a normative-rationality requirement.

The money pump argument states that those who violate transitivity stand to be exploited. Take the case where an agent has the preferences $A \succ B$, $B \succ C$, and $C \succ A$, (i.e., transitivity is violated), and she currently possesses C. Then, a cunning trader could offer to exchange (for a small fee) C for B, which the agent will accept given her preference. The trader could then do the same by offering first A and then C. In the end, the agent is left with the same alternative, C, as before the trade but has paid three small fees to the trader. Violating transitivity thus is considered irrational because it makes the violator subject to exploitation.

Another argument for the normative appropriateness of preference transitivity suggests that transitivity is constitutive of the meaning of preference. For example, Donald Davidson has claimed that violating transitivity undermines the very meaning of preferring one alternative to others.

All the arguments for preference transitivity are philosophically controversial. The money pump argument applies straightforwardly only to transitivity violations of strict, but not weak, preferences. Even then, additional conditions must be satisfied for the cunning trader to be able to arrange the money pump. A critic of the meaning argument can point out possible connections between intransitive preferences and choice.

Regardless of their substantial legitimacy, transitivity and completeness are—among others—technical prerequisites for the numerical representation of preferences through a utility function. Various axiomatic systems exist in which such representation theorems are proven (as in Decision Theory). Crucially, these utility functions not only represent the preference ordering but also allow the

measurement of preference intervals—that is, how much more an alternative *A* is preferred to *B* than another alternative *C* is preferred to *D*.

In the social sciences, preferences are commonly related to choice. We can represent choice as a function ***C*** on the set of alternatives ***X***. ***C(X)*** then delivers a set of "chosen elements" from ***X***. Economists standardly define preference in terms of choice. According to the revealed preference approach, *A* is (weakly) preferred to *B* if and only if *A* is chosen from a set of alternatives when *B* was available. Formally,

$$A \succcurlyeq^R B \text{ if and only if for some } \mathbf{X}, \quad A \in \mathbf{C(X)} \text{ and } B \in \mathbf{X}.$$

If the choice function is defined over all subsets of ***X***, $\succcurlyeq^R$ is complete. However, $\succcurlyeq^R$ does not necessarily satisfy transitivity. To secure this, various constraints have to be imposed on the choice function ***C(X)***.

There is considerable disagreement over how to interpret this relationship between preferences and choice. In its strongest form, revealed preference theorists sometimes hold that preference is nothing but choice—just a convenient shorthand, really. Against such a strong behaviorist position, philosophers have argued that choices and preferences are in fact entities of quite different categories. Preferences are states of mind, whereas choices are actions. Some, like Daniel Hausman, further argue for a broadly folk-psychological notion of preferences, where preferences in combination with beliefs cause choice.

Many economists, however, reject this folk notion as the "causal utility fallacy," as explained by Ken Binmore (2009): In revealed preference theory, it is not true that an agent chooses *B* rather than *A* because she prefers *B* to *A*. "On the contrary, it is because [the agent] chooses *B* rather than *A* that we say [the agent] prefers *B* to *A*, and assign *B* a larger utility" (p. 19). This perspective does not imply that preferences are identical to choice, but it shows that economists measure something with the revealed preference approach that considerably differs from the folk-psychological concept. What exactly the nature of this measured concept is needs to be further explored.

Till Grüne-Yanoff

See also Allais Paradox; Cost–Benefit Analysis; Decision Theory; Game-Theoretic Modeling; Homo Economicus; Judgment Aggregation and the Discursive Dilemma; Rational Choice and Political Science; Rational Expectations; Rationality and Social Explanation; Social Choice Theory; Welfare Economics

Further Readings

Anand, P. (1993). The philosophy of intransitive preference. *Economic Journal, 103,* 337–346.

Binmore, K. (2009). *Rational decisions*. Princeton, NJ: Princeton University Press.

Chang, R. (Ed.). (1997). *Incommensurability, incomparability, and practical reason*. Cambridge, England: Cambridge University Press.

Davidson, D. (1980). Hempel on explaining action. In *Essays on actions and events* (pp. 261–275). Oxford: England: Oxford University Press.

Davidson, D., McKinsey, J., & Suppes, P. (1955). Outlines of a formal theory of value, I. *Philosophy of Science, 22,* 140–160.

Grüne-Yanoff, T. (2007). Bounded rationality. *Philosophy Compass, 2*(3), 534–563.

Hausman, D. (2011). *Preference, value, choice, and welfare*. Cambridge, England: Cambridge University Press.

Samuelson, P. (1938). A note on the pure theory of consumers' behaviour. *Economica, 5,* 61–71.

Prejudice and Stereotyping

Prejudice and stereotyping, which represent two different facets of social bias, have profound impact for social justice and hierarchy. *Prejudice* is an individual-level attitude (whether subjectively positive or negative) toward groups and their members that creates or maintains hierarchical status relations between groups. *Stereotypes* are associations and beliefs about the characteristics and attributes of a group and its members that shape how people think about and respond to the group.

This entry explains each notion separately as well as how they operate jointly and shows the importance of both for understanding intergroup bias and its effects on how groups are structured hierarchically vis-à-vis each other. This pair of notions therefore is important in social-scientific inquiries about social cognition and group identity.

Prejudice

Psychologists have traditionally conceptualized prejudice as a negative attitude (i.e., a general evaluative appraisal) characterized by "antipathy." Recent definitions, however, have focused more on the functional aspects of prejudice than on the general negativity associated with it. Prejudice functions

psychologically by subjectively organizing people's environment and orienting them to objects and others within it. Prejudice also serves other purposes, such as enhancing self-esteem, maintaining group hierarchy (social stratification), and providing material advantage.

Recent definitions of prejudice bridge the individual-level emphasis of psychology and the group-level focus of sociology by concentrating on the dynamic nature of prejudice. Prejudice operates through individuals' reactions to maintain status and role differences between groups in ways that produce and reinforce structural disadvantage for another group. People who deviate from their group's traditional role arouse negative reactions; others who exhibit behaviors that reinforce the status quo elicit positive responses. Consistent with this view, prejudice toward women has both "hostile" and "benevolent" components. Hostile sexism punishes women who deviate from a traditional subordinate role, whereas benevolent sexism celebrates women's supportive, but still subordinate, position. Thus, current prejudices do not always include only an easily identifiable negative view about the target group but may also include more subtle, but patronizing and also pernicious, "positive" views.

In addition, people are not necessarily aware of their prejudices. Although prejudices may be explicit (controllable and deliberate), they may also be implicit (automatically activated and unconscious). Whereas explicit prejudice is assessed with conventional self-report measures, implicit prejudice is typically measured with response latency measures, such as the Implicit Associations Test.

Stereotyping

Early research conceived of stereotyping as a faulty thought process and focused on the content of the stereotypes of different groups. The current social-cognitive approach places more emphasis on the dynamics of stereotyping and, as with prejudice, its functions. Stereotypes guide the way information is attended to, processed, and recalled. Stereotypes not only reflect beliefs about the traits that characterize the typical member of a group and distinguish the group from others but also contain information about other qualities, such as social roles. They imply a substantial amount of other information about a person besides what is immediately apparent and generate expectations about individual group members beyond the current situation. Like prejudice, stereotyping frequently occurs automatically and implicitly.

In general, stereotypes produce a readiness to perceive behaviors or characteristics associated with the stereotype. At the earliest stages of perceptual processing, stereotype-consistent characteristics are attended to most quickly. People also recall stereotype-consistent information better than stereotype-inconsistent or stereotype-irrelevant information. Nevertheless, when there are cues in the immediate environment that prime these memories or when people are motivated to form accurate impressions of others, stereotype-inconsistent information is better recalled than stereotype-consistent information.

Stereotypes are particularly rigid and resistant to change because they distort perceptions and interpretations in ways that reinforce the stereotype. Stereotypes also justify discrimination. People infer the characteristics of groups based on the social roles that they appear to occupy. As a consequence, groups that have lower socioeconomic status (which may be caused by discrimination) are assumed to be less competent and/or less motivated than are members of groups of higher status. Minority group members may also be socialized to adopt these system-justifying ideologies.

Stereotyping *and* Prejudice

Although prejudices toward and stereotypes about different groups have unique characteristics related to specific intergroup histories or current circumstances, general principles also shape the content of stereotypes cross-culturally. The "Stereotype Content Model" proposes that the nature of stereotypes is determined by perceptions of where a group stands on two dimensions: warmth and competence. The resulting stereotypes arouse specific emotional reactions to a group, which in turn motivate actions toward the group.

Groups low on both warmth and competence (e.g., undocumented immigrants and poor people) are associated with feelings of disgust, anger, and resentment. These are particularly dehumanizing emotions; they lead to both passive harm (e.g., neglect) and active harm (e.g., attack). Groups high

in warmth but low in competence (e.g., housewives, the elderly) produce pity and sympathy. Pity elicits both passive neglect and social isolation, simultaneously with active help but often in paternalistic forms. Groups low in warmth but high in competence (e.g., Asians, Jews) elicit envy and jealousy. People are distrustful of these groups and react with pleasure to their misfortunes (*schadenfreude*). Finally, groups high in warmth and high in competence (e.g., the in-group, close allies) elicit pride and admiration, which in turn motivate positive actions, including both active and passive help.

Conclusion

Prejudice and stereotyping are two different forms of intergroup bias. Prejudice reflects *general* evaluative and affective orientations toward a group; stereotyping reflects *specific* characteristics ascribed to a group and its members. Both operate implicitly as well as explicitly. These biases originate from particular historical relations between groups and from current structural relations between groups (along the dimensions of warmth and competence). Stereotyping and prejudice both function, frequently in coordinated ways, to establish, reinforce, and justify hierarchical relations between groups.

John F. Dovidio

See also Feminist Critiques of Social Science Applications; Group Identity; Implicit Bias and Social Cognition; Racial Critiques of Social Science Applications; Social Cognition; Unconscious Social Behavior

Further Readings

Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (Eds.). (2010). *SAGE handbook of prejudice, stereotyping, and discrimination*. London, England: Sage.

Fiske, S. T. (2011). *Envy up, scorn down: How status divides us*. New York, NY: Russell Sage Foundation.

Gendler, T. (2011). On the epistemic costs of implicit bias. *Philosophical Studies, 155*(1), 33–63.

Glick, P., & Fiske, S. T. (2001). Ambivalent sexism. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 33, pp. 115–188). San Diego, CA: Academic Press.

Primatology and Social Science Applications

Why should students of the human mind and human society have any interest in nonhuman primates? This entry answers this question by putting forward three important applications of primate research to human sciences. First, understanding the behavior of our closest biological relatives can help us understand the evolutionary origins of human behavior. Second, research on related primate species can help us understand the mechanisms underlying human behavior. Finally, a comparison of various solutions that other primate species have found for coping with social life may inform us of alternative ways to think about and, possibly, alter our own behavior.

There are two major ways to search for evolutionary similarities. The term *homologies* refers to traits we have in common with our closest evolutionary ancestors, the great apes. Shared traits among related species indicate that these traits must have been shared by our common ancestors. The further back in phylogeny that one can find shared traits, the older the evolutionary history of such traits. For example, many species, including macaques, quail, and chinchillas, categorize human speech sounds in a similar way, suggesting that we have constructed our language on ancient perceptual structures. Homologies represent a shared evolutionary heritage, and departures from homologies indicate evolutionary divergence.

The term *homoplasies* refers to traits that are not shared with common ancestors but are shared instead with species that appear to have similar adaptive problems, with similar structures or behaviors to adjust to these problems. Thus, whereas we share many of our cognitive and physical traits with chimpanzees and bonobos, our closest relatives, our pair-bonded mating system with shared parental care is not seen in any great ape species but can be seen in pair-bonded primates that are phylogenetically removed from humans. These species also share with humans a relatively rich system of vocal communication compared with great apes, as well as greater sensitivity to others' needs. The convergence of these traits in pair-bonded monkeys and humans suggests an evolutionary convergence that is likely due to common solutions to common adaptive problems.

The origins of human behavior are therefore a mixture of homologies and homoplasies, and only by examining a divergent range of nonhuman primate species can we develop a full picture of human evolution. Some illustrative examples are as follows.

We can see in chimpanzees the origins of some of our cognitive skills in the creation and use of tools, the rapid learning and use of arbitrary symbols as words for communication, mirror recognition, the acquisition of numerosity skills, and the complexity of memory. However, chimpanzees appear relatively unable to share (especially in the context of food), to teach offspring, or to learn socially from others (except in competitive contexts). However, rapid social learning, sharing of food resources (even at a cost to the sharer), cooperation, and even rudimentary teaching behavior are seen in cooperatively breeding primates quite distant in evolution from humans. The primatologist Sarah Hrdy has argued that humans are cooperative breeders and that our cognitive skills are the result of an amalgam of cognitive skills shown by chimpanzees with the social sensitivity and other-regarding abilities of cooperative breeding primates.

Human social organization is fundamentally different from that seen in any single species of nonhuman primate, but the primatologist Bernard Chapais has argued that a synthesis of information from pair-bonded primates evolutionarily far removed from the great apes, where fathers play an active role in infant caretaking, coupled with the male philopatry seen in apes, has led to bilateral kinship relationships with the kin of both parents, leading to the emergence of tribes, which characterize early human social groups. Again, the combination of homology and homoplasy provides an accurate understanding of how humans evolved.

The emergence of morality is another area where an understanding of multiple primate species can inform the origins of human behavior. Many studies have looked at reconciliation following an aggressive act as a mechanism to restore social homeostasis within a group. Evidence from several species also suggests that primates evaluate equity, actively rejecting food (which would have been preferred otherwise) if another receives a greater amount. Primates also behave reciprocally with respect to cooperating to obtain resources. In some species, we see mechanisms to punish those who fail to follow rules. At the same time, one can find the emergence of empathy, grief, and other emotional reactions often attributed solely to humans—as is shown in studies by the primatologist Frans de Waal.

Specific types of sounds (harmony versus dissonance, staccato versus legato notes, and broad-band versus narrow-band sounds) appear to convey or induce emotions and are used by humans with both infants and pets. Constructing music in the frequency range and tempos of other species with these principles can induce either calming or aroused behavior depending on the structures used. These findings suggest that music may have emerged from emotional communication and that the emotional components of music have a long evolutionary history.

Developmental, behavioral, and neural mechanisms affecting behavior may be more clearly studied in nonhuman species. One can study in primates the development of social, cognitive, and communication skills under controlled environments. Communication skills, fear of predators, and parenting skills emerge through learning and social interactions. Mirror neurons that may form the basis for imitation were first discovered in monkeys and were inferred to exist in humans. The hormonal basis for pair bonding and male care of infants was documented first in pair-bonded primates and subsequently extended to humans.

Finally, understanding the diversity of behavioral solutions found in nonhuman primates might suggest novel solutions to human problems. There is no evidence of dominance or aggression between or within the sexes in muriqui monkeys in Brazil. Learning how they manage social life peacefully may have relevance to humans. Since pair bonding and biparental care are key components of human social cognition and social organization, knowledge of the social organization, behavior, and biology of nonhuman primates that also exhibit pair bonding and biparental care is of critical importance.

Charles T. Snowdon

See also Biology and the Social Sciences; Cooperation, Cultural Evolution of; Evolutionary Ethics; Evolutionary Game Theory and Sociality; Evolutionary Psychology; Human Cultural Niche Construction and the Social Sciences; Mirror Neurons and Motor Cognition in Action Explanation; Naturalism in Social Science; Sociobiology

Further Readings

Chapais, B. (2008). *Primeval kinship*. Cambridge, MA: Harvard University Press.

De Waal, F. (2006). *Primates and philosophers: How morality evolved* (S. Macedo & J. Ober, Eds.). Princeton: NJ: Princeton University Press.

De Waal, F. (2009). *The age of empathy.* New York, NY: Harmony Books.

Hrdy, S. B. (2009). *Mothers and others: The evolutionary origins of mutual understanding.* Cambridge, MA: Belknap Press.

Snowdon, C. T. (2009). Plasticity of communication in nonhuman primates. *Advances in the Study of Behavior, 40,* 239–276.

Snowdon, C. T., Pieper, B. A, Boe, C. Y., Cronin, K. A., Kurian, A. V., & Ziegler, T. E. (2010). Variation in oxytocin levels is associated with variation in affiliative behavior in monogamous pairbonded tamarins. *Hormones and Behavior, 58*(4), 614–618.

Snowdon, C. T., & Teie, D. (2010). Affective responses in tamarins elicited by species-specific music. *Biology Letters, 6,* 30–32.

Tomasello, M. (2008). *Origins of human communication*. Cambridge: MIT Press.

Tomasello, M., & Call, J. (1997). *Primate cognition*. Oxford, England: Oxford University Press.

Probability

Probability plays several crucial roles in science: Analysis of statistical data, modeling of both physical and social systems (e.g., rational choice models in economics), and risk analysis would all be impossible without it. The subject has *mathematical*, *metaphysical*, and *epistemological* aspects. This entry reviews these three aspects in turn.

Mathematical Aspects of Probability

The mathematical foundations of probability are simple. When analyzing some scenario probabilistically, we select a space of *basic possibilities*, requiring that they be exhaustive (at least one must obtain) and exclusive (no more than one can obtain)—for example, in a toss of a pair of dice, the 36 possible ways the dice can land are usually taken to be the basic possibilities. Each possibility gets assigned a number in the range (0,1), which represents its *probability*—1/36 in our example, assuming that the dice are fair and independent. A *proposition* (what mathematicians call an "event") is something that turns out true according to some of the basic possibilities and false according to the rest (e.g., "at least one die lands 6"). The probability of any proposition is then simply the sum of the probabilities of those possibilities that make it true.

Standard axioms for probability theory fall out immediately: Given any propositions A and B,

i. $0 \leq \text{Prob}(A) \leq 1$;

ii. if A is a *contradiction* (true according to no possibilities), then $\text{Prob}(A) = 0$;

iii. if A is a *tautology* (true according to all possibilities), then $\text{Prob}(A) = 1$; and

iv. if A and B are *incompatible* (no possibility makes both true), then $\text{Prob}(A \text{ or } B) = \text{Prob}(A) + \text{Prob}(B)$. (Some presentations use an infinitary version of this axiom.)

It is then usual to define the *conditional probability* of A, given B, as

$\text{Prob}(A, \text{given } B) =_{df} \text{Prob}(A \text{ and B})/\text{Prob}(B)$, when $\text{Prob}(B) \neq 0$.

A and B are *independent* iff $\text{Prob}(A \text{ and } B) = \text{Prob}(A) \bullet \text{Prob}(B)$. If so, then $\text{Prob}(A, \text{given } B) = \text{Prob}(A)$ when the left-hand side is defined.

This presentation simplifies one detail, since we have assumed that the space of possible worlds is at most countably infinite; if it is uncountably infinite, then, typically, each basic possibility gets a probability of zero, and we apply techniques from calculus to integrate over sets of basic possibilities in order to arrive at nonzero probabilities for propositions. The axioms listed above still hold.

Beyond laying down the requirement that the basic possibilities be exclusive and exhaustive, the mathematical structure of probability does not dictate the choice of what they are, let alone what their probabilities are. In practice, some judgment is required in making this choice, so that subsequent analysis can proceed in the clearest and most illuminating fashion.

Metaphysical Aspects of Probability

Turn now to metaphysics: What is a statement of probability *about*? For example, if a meteorologist

asserts that the probability that it will rain tomorrow is 30%, what must our world be like for that assertion to be *true*? Approaches to this question display a striking disparity.

On the *classical* approach—now fallen into disfavor—the probability of a proposition is simply the ratio of those basic possibilities according to which it is true to the total number of basic possibilities. This approach assumes, implausibly, that in any given application, there is a uniquely best way to select the set of basic possibilities and that each possibility in the set should be treated as equally probable.

According to *subjectivist* approaches, statements of probability are facts about the *degrees of confidence* rational agents do or should have in the proposition in question. Rational opinion can come in degrees intermediate between full belief and full disbelief; the subjectivist takes the axioms of probability to articulate the structure these degrees of confidence ought to display. A subjectivist might say that our meteorologist is asserting that her own degree of confidence that it will rain tomorrow, in light of all the evidence available to her, is 0.3.

Relatedly, the *logical* approach to probability claims that an assertion of probability is an assertion about the *degree of logical support* certain premises (often tacitly specified) give to some conclusion. Thus, our meteorologist's claim is true just in case the total evidence available to her logically supports the conclusion that it will rain tomorrow to degree 0.3. Spelling out the details of this notion of logical support remains an elusive goal for fans of this approach.

Objectivist approaches treat probabilities as physical features of the world—that is, independently of how rational agents do or ought to think. Thus, frequentism claims, roughly, that statements about probability are statements about long-run frequencies; our meteorologist might be understood as asserting that in 30% of those situations that are, in relevant meteorological respects, like the present one, rain follows a day later. A persistent difficulty is to say in detail how facts about frequencies constitute facts about probabilities—particularly given that, in actual practice, we often take it for granted that frequencies can fail to match the underlying probabilities (e.g., it is surely possible for a fair coin to land heads every time it is tossed.). One way to preserve objectivism in the face of the difficulties that beset frequentism is to go primitivist and insist that there is no way to reduce facts about objective probabilities to any other sorts of facts (concerning frequencies or anything else).

One need not choose between subjectivism and objectivism: In fact, it has become common for authors to argue that there are different kinds of probability, for some of which a subjectivist approach works and for others, an objectivist approach. Important recent work takes pluralism for granted and seeks to articulate precise principles connecting different species of probability.

Epistemological Aspects of Probability

That there is no settled view about the metaphysics of probability is important to keep in mind when considering how to *establish* or *test* a claim concerning probabilities: The relevant kinds of evidence or argument may depend on whether the claim is interpreted in a subjectivist or objectivist manner. Regardless, probably the most important principle in the epistemology of probability is that one should expect probabilities to be reflected, with reasonable accuracy, in the long-run statistics of suitably designed experiments. In the ideal case, one runs an extremely large number of repetitions of the same experiment in such a way that one can be confident that the outcomes of distinct trials are *independent* of one another and that the probability of each kind of outcome remains *constant* from trial to trial. Then, one can prove that the long-run frequencies will, with high probability, accurately reflect these underlying probabilities (this is the *weak law of large numbers*).

But two caveats complicate the epistemology of probability considerably. First, it is only rarely, in actual scientific practice, that one can meet these conditions. Second, emphasizing statistics of repeated experiments can blind one to other sources of knowledge about probability. An illustrative example follows: Presented with a coin of unknown bias, you may have no way of determining this bias without tossing the coin repeatedly. For all that, you can know, on general physical grounds, that the probability that a pair of tosses will land heads, then tails, is exactly the same as the probability that it will land tails, then heads. That is a fact about the probabilistic behavior of the coin, ascertainable without the need of performing repeated pairs of tosses and observing their statistics.

Ned Hall

See also Bayesianism, Recent Uses of; Decision Theory; Formal Epistemology; Induction and Confirmation; Econometrics: Methodological Issues

Further Readings

Jeffrey, R. (1990). *The logic of decision.* Chicago, IL: University of Chicago Press.

Lewis, D. (1986). A subjectivist's guide to objective chance. In *Philosophical papers* (Vol. 2, pp. 83–113). Oxford, England: Oxford University Press.

Mellor, D. H. (2005). *Probability: A philosophical introduction.* London, England: Routledge.

Promises and Agreements

Of the many obligations that we acquire in our lives, those that arise from promises and agreements are among the most commonplace. Any general theory of moral obligation ought therefore to be able to explain what promises are and why they obligate. Modern moral philosophy, much concerned with ideas of duty and obligation, has thus been impatient to provide such an explanation. But the humble promise has been stubborn in refusing to give up its secrets. Our ability to give another the right to demand that we perform some act that was previously wholly optional, just by uttering a few inoffensive, prosaic words, has appeared opaque even under the lights of some of the most penetrating philosophical minds. "Since every new promise imposes a new obligation," David Hume was driven to remark in his *Treatise on Human Nature* (1739/2000), "'tis one of the most mysterious and incomprehensible operations that can possibly be imagin'd, and may even be compar'd to *transubstantiation* or *holy orders*" (p. 336). A theory of promise is a theory that explains what a promise is and how it creates obligations.

This entry surveys the prevailing theories of promise and concludes with some remarks on the relationship between promises and agreements. Recent philosophy of the social sciences has developed elaborate views on collective agency and collective intentionality, placing them at the center of philosophical theories of sociality. Such shared acts and collective intentions are based on notions such as joint commitment, promising, agreement, mutual beliefs or mutual obligations, and the like; it is therefore important for such a social ontology to be clear about what agreeing and promising amount to. In addition, these two ideas were important to historical social contract theories that have also been central to philosophical theories about the constitution or creation of social institutions based on joint commitment and social conventions. Promises and agreements are such vehicles of creating sociality.

A promise is a communicative act, and as such it can be analyzed in terms of the communicative intentions of a promisor. Promising is the act of uttering "I will φ" (or an equivalent future affirmation), with the intention of being understood to be promising to φ. Theories of promise can be distinguished according to the way they characterize this intention. Most fall into one of two camps.

Theories in the first camp conceive of promising as a creative act whereby a speaker communicates an intention to create, by that act of communication, an obligation to the hearer to do what he is saying that he will do. This "creative-act" conception of promising finds its origins in Hume and has been advanced in more recent times by philosophers and legal theorists like Elizabeth Anscombe, John Finnis, H. L. A. Hart, David Owens, John Rawls, Joseph Raz, Seana Shiffrin, and John Searle, among others.

The creative-act conception of promising presupposes a rule of promise keeping that renders this morally inert speech act normatively significant. Many creative-act theorists hold that this rule is embedded in a social practice the violation of which contravenes some moral principle, such as one forbidding people from departing from just social practices. Others insist that the obligation of a promise may be institutional or conventional without also being moral. Still others hold that the norm of promise keeping is a moral one that can be justified without resort to the idea of a social practice.

Theories in the second camp hold that what is essential to promising is not that the speaker communicates an intention to become obligated but that he communicates that he intends to φ, with the aim of assuring the hearer that he will φ. These theories begin from the observation that the ordinary pragmatic function of a promise is to assure the hearer that the speaker will adhere to some course of conduct and to count on that belief in her deliberations and planning. They conceive of promises as (or as involving) ordinary (albeit often emphatic)

statements of intention made with a view to inspiring assurance, reliance, trust, or the like in the hearer.

These "expectational" theories of promise seek to explain the obligation of promises without invoking a special rule of promise keeping. They hold that breaking a promise is wrong, when it is wrong, because it amounts to a kind of mistreatment of the sort that is proscribed by some moral principle of general application, such as that which requires that we not betray the trust we invite others to place in us or that which forbids the doing of avoidable harm to others. The idea is that once a speaker has invited the hearer to form and possibly act upon a belief that the speaker will follow some course of conduct φ that he knows the hearer wants to be assured of, his failure to φ is a failure to treat the hearer with due respect. Some early utilitarians, such as Adam Smith and John Stuart Mill, subscribed to an expectational view of promising. In more recent times, expectational theories have been advanced by Páll Árdal, Patrick Atiyah, Neil MacCormick, Thomas Pink, and Thomas Scanlon, among others.

These, then, are the contours of the two main schools of thought in the philosophy of promising. What of agreements? In contrast to promises, which are the focus of a prodigious literature, the equally pervasive social phenomenon of agreements has attracted relatively little philosophical attention. This neglect may be the result of the widely held assumption that agreements are just exchanges of promises. This assumption is implicit in everyday discourse, where "promise" and "agree" are often used interchangeably and people who dishonor agreements are censured for breaking their promises; and it is orthodoxy in the law that defines contracts as enforceable agreements and agreements as exchanges of promises.

In a 1993 article titled "Is an Agreement an Exchange of Promises?" Margaret Gilbert challenged this assumption. Gilbert argued that no two promises, when exchanged, will yield obligations for the promisors that are both interdependent and simultaneous. Since interdependence and simultaneity are essential features of the obligations that are acquired by the parties to an agreement, moreover, an agreement is not a promise exchange. Gilbert's article has spawned a small but growing literature on the kinship between promises and agreements, including work by Kent Bach, Oliver Black, and Hanoch Sheinman.

Michael Pratt

See also Collective Agents; Collective Intentionality; Collective Moral Responsibility; Disagreement; Law, Social Phenomenon of; Plural Subjects; Social Contract Theories; Social Conventions; Social Institutions; Social Rules; Speech Acts; Trust, Social

Further Readings

Anscombe, G. E. M. (1978). Rules, rights, and promises. *Midwest Studies in Philosophy, 3,* 318–323.

Árdal, P. (1968). And that's a promise. *Philosophical Quarterly, 18,* 225–237.

Atiyah, P. (1981). *Promises, morals and law.* Oxford, England: Clarendon Press.

Bach, K. (1995). Terms of agreement. *Ethics, 105,* 604–612.

Black, O. (2004). Agreements, undertakings, and practical reason. *Legal Theory, 10,* 77–95.

Finnis, J. (1980). *Natural law and natural rights.* Oxford, England: Clarendon Press.

Gilbert, M. (1993). Is an agreement an exchange of promises? *Journal of Philosophy, 90,* 627–649.

Hart, H. L. A. (1958). Legal and moral obligation. In A. I. Melden (Ed.), *Essays in moral philosophy* (pp. 82–107). Seattle: University of Washington Press.

Hume, D. (2000). *A treatise of human nature* (D. Norton & M. Norton, Eds.). Oxford, England: Oxford University Press. (Original work published 1739)

MacCormick, N. (1972). Voluntary obligations and normative powers. *Proceedings of the Aristotelian Society, 46*(Suppl.), 59–78.

Mill, J. S. (1991). *On liberty.* Oxford, England: Oxford University Press. (Original work published 1859)

Owens, D. (2006). A simple theory of promising. *Philosophical Reviews, 115,* 51–77.

Pink, T. (2009). Promising and obligation. *Philosophical Perspectives, 23,* 389–420.

Rawls, J. (1972). *A theory of justice.* Oxford, England: Oxford University Press.

Raz, J. (1977). Promises and obligations. In P. Hacker & J. Raz (Eds.), *Law, morality and society* (pp. 210–228). Oxford, England: Clarendon Press.

Scanlon, T. M. (1998). *What we owe to each other.* Cambridge, MA: Belknap.

Searle, J. (1969). *Speech acts: An essay in the philosophy of language.* Cambridge, England: Cambridge University Press.

Sheinman, H. (2011). Agreement as joint promise. In *Promises and agreements: Philosophical essays* (pp. 365–395). New York, NY: Oxford University Press.
Shiffrin, S. (2008). Promising, intimate relationships, and conventionalism. *Philosophical Review, 117,* 481–524.
Smith, A. (1978). *Lectures on jurisprudence* (R. Meek, D. D. Raphael, & P. Stein, Eds.). Oxford, England: Clarendon.

Prophecy, Self-Fulfilling/Self-Defeating

Robert Merton (1968) coined the phrase *self-fulfilling prophecy* to describe "a *false* definition of the situation evoking a new behavior which makes the originally false conception come *true*" (p. 477). It is not the truism that people's perceptions depend on their prior beliefs. Nor is it the truism that beliefs, even false ones, have real consequences. To count as self-fulfilling prophecy, a belief must have consequences of a peculiar kind: consequences that make reality conform to the initial belief. The converse is the *self-defeating prophecy*, where the initial belief leads to behavior that falsifies the belief.

The latter has attracted far less attention from social scientists, so this entry will focus on the self-fulfilling kind. The entry begins with a definition and then discusses some plausible instances.

A simple example will aid definition. First, a teacher believes that a particular student has exceptional academic ability. Second, this belief leads the teacher to encourage and stimulate the student. Third, the teacher's behavior causes the student to manifest exceptional academic ability. A self-fulfilling prophecy implies that any change in Step 1 would alter the outcome: If the teacher had believed the same student to be mediocre, then the teacher would have behaved very differently, and in consequence, the student would have performed poorly. This counterfactual assumption avoids the problem of specifying the prior reality. There is no need to decide whether the student was really exceptional or mediocre at the outset; what matters is that altering the teacher's belief would transform the outcome accordingly. A self-fulfilling prophecy implies that the actors (or at least some of them) misapprehend the causal process by overlooking the third step. The teacher assumes that the student's presumed ability—rather than the teacher's behavior—caused the subsequent high achievement. If the true causal sequence could be recognized, then the process would be transformed. For example, the teacher would behave like this with every student, thus elevating them all.

Instances of self-fulfilling prophecies have been discerned at a range of scales. The relation between teacher and student can be extended to relationships between two collectivities. Police believe that a crowd is violent and so launch a preemptive attack; the crowd then reacts by fighting back. At a purely individual level, the placebo response can be considered a self-fulfilling prophecy. A patient believes that an inert substance is a powerful painkiller, and therefore, she feels less pain. This phenomenon is sociological as well as biological, as the patient's belief derives from a relationship with someone assumed to possess medical expertise.

The most interesting self-fulfilling prophecies involve more complex social processes, such as financial markets. Merton introduced his concept with a fictional story of a bank run. Depositors believe that the bank is insolvent and therefore withdraw their money. This sudden loss of funds drives the bank out of business. The bank run is typically used to exemplify self-fulfilling prophecies, though there is more empirical evidence for the opposite phenomenon: investment bubbles. Investors wrongly believe that the investment generates high returns and therefore invest their funds. As more people invest, the investment becomes more valuable. Moreover, subsequent investment enables a Ponzi scheme to provide initial investors with the promised returns; visible payouts further enhance the scheme's credibility. All this leads to further investment. The process can last for some time but eventually proves unsustainable. The ultimate outcome is financial failure, as for the bank run. The crucial difference is that the investment bubble's outcome negates rather than fulfills the prophecy. Ponzi schemes provide a quasi-experimental demonstration of the self-fulfilling prophecy.

In the most intriguing case, social theory makes social reality in its own image. Marxism predicts that capitalism is fated to end in revolution and thus encourages its adherents to foment revolution. Neoclassical economics, which likewise conflates description and prescription, has been more

successful. Its assumption that firms maximize profits, or equivalently shareholder value, was falsified when empirically tested in the 1950s. Instead of revising the theory, neoclassical economists argued that managers should be given appropriate incentives to maximize profits, namely, stock options. Thanks in part to their argument, companies adopted this form of compensation, and maximizing shareholder value became a social norm. As a result, short-term profit maximization—as neoclassical theory predicts—is more true today.

A self-fulfilling prophecy is not easy to demonstrate empirically. Only experimental intervention can ensure that actors' beliefs are not influenced by prior reality. To return to the example used at the outset, the investigator can select a few students at random and convince their teacher that they have been diagnosed with exceptional ability. Their subsequent academic achievement should be greater than average, if a self-fulfilling prophecy is to be operative. Such experimentation is limited by the difficulty of manipulating beliefs in human beings, especially when investigation is bound by informed consent. Experimental results will inevitably underestimate the causal impact of belief.

Self-fulfilling prophecies have an abiding fascination because they show how we can be caught in a web of our own making. By reifying social reality, we can fail to understand that we have been responsible for creating it.

Michael Biggs

See also Feedback Mechanisms and Self-Regulatory Processes in the Social Sciences; Retrodiction and the Epistemology of Future Studies

Further Readings

Biggs, M. (2009). Self-fulfilling prophecies. In P. Bearman & P. Hedström (Eds.), *The Oxford handbook of analytical sociology* (pp. 294–314). Oxford, England: Oxford University Press.

MacKenzie, D. A. (2006). *An engine, not a camera: How financial models shape markets*. Cambridge: MIT Press.

Merton, R. K. (1968). The self-fulfilling prophecy. In *Social theory and social structure* (2nd ed., pp. 475–490). New York, NY: Free Press. (Original work published 1948)

Salganik, M. J., & Watts, D. J. (2008). Leading the herd astray: An experimental approach to self-fulfilling prophecies in an artificial cultural market. *Social Psychology Quarterly, 71,* 338–355.

Pseudoscience

The term *pseudoscience* refers to a highly heterogeneous set of practices, beliefs, and claims sharing the property of appearing to be scientific when in fact they contradict either scientific findings or the methods by which science proceeds. Classic examples of pseudoscience include astrology, parapsychology, and ufology; more recent entries are the denial of a causal link between the HIV virus and AIDS or the claim that vaccines cause autism. To distinguish between science and pseudoscience is part of what the philosopher Karl Popper referred to as the *demarcation problem*, a project that has been dismissed by another philosopher, Larry Laudan, but that keeps gathering much interest in philosophers, scientists, educators, and policymakers.

This entry provides the basics of the debate about demarcation, as well as a brief discussion of why it is of vital importance not just intellectually but for society at large.

Popper and the Demarcation Problem

Popper began working on the problem of demarcation between science and pseudoscience (as well as more generally nonscience) as early as 1919. He was particularly concerned with David Hume's famous problem of induction, the idea that there does not seem to be a logically independent way to justify inductive reasoning, the basis for the scientific method. Popper thought he had arrived at a single idea that represented both a solution to Hume's problem as well as a clear-cut demarcation criterion: *falsificationism*. He proposed that science is in the business of advancing falsifiable (i.e., refutable in principle) theories about how the world works. This appeared to have bypassed Hume's issue about induction because falsificationism can be thought of as an application of *modus tollens*, therefore relying on deductive, not inductive, reasoning. At the same time, it seemed to Popper that pseudosciences (among which he counted various schools of psychoanalysis as well as Marxist theories of history) made statements that were not falsifiable, and were thereby unscientific.

While Popper's contribution to both issues remains a fundamental starting point for any discussion of demarcation and induction, there are good reasons to believe that he was a bit too quick in

declaring victory on both fronts. This entry directs the reader to two comprehensive articles concerning the problem of induction and focuses instead on falsification as a demarcation criterion.

It is easy to show that falsification leaves much good science out and allows a significant amount of pseudoscience in. For instance, the history of science is riddled with examples of scientific hypotheses that—when first proposed—were apparently falsified by the data and yet scientists kept them alive because they seemed promising enough. The initial version of the Copernican system, with its circular planetary orbits, was doing no better empirically than the Ptolemaic system it was supposed to replace, and it was not until Kepler realized that the planets move along elliptical orbits that the theory was vindicated. Copernicus's book was published in 1543, but it was not until 1609 that Kepler put out his fundamentally revised version of the theory.

On the other side of the divide, so-called scientific creationism does make perfectly falsifiable predictions, such as that the earth is only a few 1,000 years old. These predictions have indeed been amply falsified by modern geology, physics, chemistry, and biology, and yet there does seem to be a strong sense that we should not simply consider creationism a science, even a failed one (for one thing, because of its appeal to supernatural, by definition inscrutable, forces, which are themselves outside the purview of science).

What the inadequacy of falsification in establishing a demarcation criterion hinted at was something that became progressively clearer in the decades following Popper's pioneering work: Science and pseudoscience are simply not the sort of concepts that admit of being defined by a small set of individually necessary and jointly sufficient conditions, the way classical logicians would want the task to be accomplished. However, it is important to understand that this sort of situation is not limited to the case of the science/pseudoscience demarcation. Plenty of other complex and interesting concepts are too "fuzzy" (in the technical sense of fuzzy logic) to admit of sharp boundaries and clear-cut definitions. Examples include the idea of "game" (as famously pointed out by Ludwig Wittgenstein) as well as the concept of biological species.

Take games, for instance. Clearly, we seem to know what sort of activities reasonably fall into that category (chess, soccer) and which don't (war, philosophizing), and of course, we are aware of borderline cases (e.g., "games" actually used to solve practical problems, e.g., the 3-D folding of proteins). Yet it is easy to appreciate how difficult it is to come up with a small number of criteria that sharply define what a game is. For each candidate, say, "done competitively," there are both instantiations belonging to the set of interest that fail the criterion (e.g., solitaire) and others that meet it while clearly not belonging to the set (e.g., business transactions). So in some sense, it should not really be surprising that the terms *science* and *pseudoscience* are difficult to define exactly and yet still refer to reasonably coherent types of activities that are distinct in important ways from each other.

Laudan and the Alleged Demise of the Demarcation Problem

The philosopher Larry Laudan declared the demarcation problem dead, and the concept of pseudoscience useless and pernicious, in a famous article published in 1983. Laudan pointed to the "failure" of philosophers to agree on necessary and jointly sufficient criteria for demarcation as an indication of the futility of the project. Moreover, he asserted that since *pseudoscience* is an emotionally charged word (clearly always used in dismissive terms), it does not really belong to philosophical discourse. The real issue, Laudan maintained, is to assess the epistemic warrant behind each individual claim to knowledge, regardless of whether it is made from within fundamental physics or astrology.

While Laudan's critique has been important in the history of the debate about pseudoscience, there are several counterpoints to consider, which explain why philosophical discussions of demarcation have been on the rise again during the past decade, with no sign of abating any time soon. Let us start with Laudan's last point, that epistemic warrant should be attached to specific claims, not to broad endeavors. This is much too restrictive and impractical. When a field like astrology has repeatedly, and for a long time, demonstrated its inability to make progress—due to the incoherence of its theoretical constructs (e.g., "constellations" are actually optical illusions) and its failure on empirical grounds—it seems the time has arrived to archive the whole thing as not warranting any more serious investigative efforts. Within a successful and dynamic science, on the other hand, the advice to examine each claim on its own merits makes sense precisely because that

science has established methods and background knowledge against which the epistemic warrant of any new claim can be reasonably assessed. Labeling something as pseudoscience—if called for—serves the same practical shortcut function of throwing an obviously frivolous lawsuit out of court before one invests money and time in something that has no chance of succeeding.

As for Laudan's argument that philosophers have failed at the demarcation task, and that we should therefore move on, it seems to be based on a peculiar understanding of "failure" for a philosopher. It can be reasonably argued that it is precisely through the exploration and criticism of possibilities in logical space that philosophy makes progress. Popper was wrong about having solved the problem of induction, but his attempt based on replacing inductive with deductive reasoning was a potentially good move that had to be properly explored and criticized before we could consider more sophisticated proposals. The same can be said of other areas of philosophical inquiry: Utilitarianism in ethics, as originally conceived by Jeremy Bentham and then John Stuart Mill, has not survived unscathed in the modern philosophical literature; but modern utilitarians like Peter Singer have been able to develop a much more nuanced view of their approach to moral philosophy precisely because they have been confronted with several rounds of criticism. The abandonment of the quest of necessary and jointly sufficient criteria to define science and pseudoscience in favor of, for instance, Wittgenstein-type family resemblance ("fuzzy") concepts constitutes progress, not failure.

Finally, let us consider Laudan's point about the emotional ladenness of the term *pseudoscience*. There is no doubt that this is the case, but Laudan himself argues that philosophy should be able to tell us what is reasonable to believe and what is not, and it is hard to imagine how "unreasonable" is the kind of label that would be much more palatable and emotionally neutral than "pseudoscience." Indeed, here, Laudan actually hits the nail on the head in implying that a major role of philosophy of science is to be prescriptive, particularly when it comes to discussions of science in the public sphere, where the consequences of our views are not merely academic but involve policy and politics.

The Role of Philosophy in Combating Inferential Biases

The current philosophical literature on pseudoscience is exploring some of the alternatives to the classical demarcation approach briefly mentioned above, such as solutions based on fuzzy logic or on making more precise the notion of Wittgensteinian family resemblance concepts. A significant number of papers have come out recently vigorously debating whether supernatural claims (e.g., Intelligent Design [ID] creationism) belong to a separate category of pseudoscience by virtue of their very invocation of the supernatural or whether they are simply another type of unscientific claim along the lines of astrology, ufology, and so on.

This particular discussion has, again, very tangible social repercussions, as was clearly on display during the 2005 trial over the teaching of ID in public schools in Dover, Pennsylvania. During the court proceedings there, both Barbara Forrest and Robert Pennock had the increasingly less rare distinction of being called as witnesses by virtue of being philosophers of science, and hence capable of providing expert testimony on the scientific nature (or lack thereof) of ID theory. As a result, Judge John E. Jones III's decision in the Kitzmiller versus Dover Area School District is a nice summary of the philosophical issues surrounding the demarcation problem. The judge came down on the side of philosophers, who were arguing in a prescriptive fashion that ID is a pseudoscience, one of the reasons why it should not be taught as science in the district's public schools (another reason being, as Judge Jones also noted, that ID is no different in substance from classical creationism, which other courts have found to be a religious doctrine, the teaching of which would violate the establishment clause of the First Amendment to the Constitution of the United States).

Pseudoscientific notions will likely stay with us for a long time, as recent literature in psychology shows that human beings are naturally prone to a number of cognitive biases that favor the persistence of pseudoscience. Interestingly, many of these biases find their equivalent in the philosophical literature on logical fallacies (e.g., the post hoc, ergo propter hoc fallacy—by which one confuses correlation with causation—is a major mechanism that we use to

make preliminary inferences about causality). On the positive side, psychologists have shown that an awareness of cognitive biases diminishes one's proneness to perpetuate the mistake. This in turn would seem to suggest that the teaching of critical thinking skills in philosophy classes is a fundamental component of the education of an intelligent citizenry.

Massimo Pigliucci

See also Debunking Social Science; Explanation, Theories of; Falsifiability; Popper's Philosophy of Science; Scientific Method

Further Readings

Boudry, M., Blancke, S., & Braeckman, J. (2010). Irreducible incoherence and intelligent design: A look into the conceptual toolbox of a pseudoscience. *Quarterly Review of Biology, 85*(4), 473–482.

Hansson, S. O. (2009). Cutting the Gordian knot of demarcation. *International Studies in the Philosophy of Science, 23,* 237–243.

Laudan, L. (1983). The demise of the demarcation problem. In R. S. Cohen & L. Laudan (Eds.), *Physics, philosophy and psychoanalysis* (pp. 111–129). Dordrecht, Netherlands: D. Reidel.

Okasha, S. (2001). Hume and induction. *Philosophical Quarterly, 51,* 307–327.

Pigliucci, M., & Boudry, M. (Eds.). (2012). *The philosophy of pseudoscience: Revisiting the demarcation problem.* Chicago, IL: University of Chicago Press.

Popper, K. (1957). Philosophy of science: A personal report. In C. A. Mace (Ed.), *British philosophy in the mid-century* (pp. 155–191). Crows Next, New South Wales, Australia: Allen & Unwin.

Wittgenstein, L. (1958). *Philosophical investigations.* Hoboken, NJ: Blackwell.

Psychoanalysis, Philosophical Issues in

Philosophical discussions of psychoanalysis have frequently focused on two topics: (1) how well psychoanalytic theories can be regarded as *evidentially supported* by the clinical data they are initially framed to explain and (2) how far particularly psychoanalytic conceptions of *unconscious mental states and processes* should be regarded as viable. The first of these will be the main topic of this entry, and the second will be briefly considered at the close.

Free Association and Freud's Claim About Evidence

The relevant clinical data arise in the practice of *free association*, as pursued by patients in analysis over the course of months and years. To free-associate is to describe the contents of what is sometimes called the *stream of consciousness*—passing experiences, thoughts, feelings, and so forth, as fully as possible as they occur and without omitting or censoring anything. This leads rapidly to thoughts and feelings that are unexpected even to those thinking them.

The data thus generated enabled Sigmund Freud and his successors to learn as much about what went on in the minds of their patients as the patients were able to put into words and, in addition, to base further conclusions on patterns that could be observed while they were doing so. Thus, there were the patterns relating associations to elements of the manifest contents of dreams, described by Freud in his own case in *The Interpretation of Dreams*. Again, there was the pattern Freud described as *transference*, in which emotions and conflicts felt early in life toward parents and siblings were revived in patients' current experience of the therapist. Since free association and the experiences that emerged in analysis were Freud's main sources of data, he maintained that persons who did not have firsthand experience of them were not in a position to criticize his theoretical conclusions.

Dispute About Freud's Claim: Advocates and a Comparison With Darwin

This claim has been at the core of disputes about evidence in psychoanalysis. Advocates characteristically maintain that the claim reflects the fact that psychoanalysis has a unique and remarkably rich source of data that are otherwise unfamiliar and unexpected. Such data cannot be ignored, but they also cannot readily be communicated, except in small and isolated vignettes. Freud's description of the publishable parts of his associations to elements

of his dream of Irma's Injection, for example, takes up many pages; and even such a detailed written description of mental processes leaves out much of the information available to the subject experiencing the processes at firsthand. When we reflect that such an accumulation of data is relevant for just a partial understanding of a single dream, we can see that there is no prospect of adequately summarizing or surveying the full range of data relevant to the main conclusions to be drawn from even a relatively short period of interpretive thinking of this kind.

In this perspective, we might compare psychoanalysis with Darwin's theory of natural selection. Darwin's theory also relates to a great range of data, drawn both from the observation of current forms of life and from countless traces of extinct and very different forms left in the incomplete and largely uncharted fossil record. Here again, we have a vast field of potential data, much of which remains unknown. These data, however, are relatively enduring and publicly available for classification, discussion, and survey. So investigators working with them have been able to use them to show a larger and often skeptical scientific public how they confirm the kind of theories Darwin and his successors devised. In the case of psychoanalysis, by contrast, the data are private and perishing and emerge in each analysis in ways unique to the individuals concerned. So for psychoanalysis, despite a comparable wealth of data drawn from countless hours of individual analyses, there can be no such disciplined public demonstration of confirmation, or any such assuaging of skepticism, as obtains for Darwin's theory.

Critics and the Questioning of Evidential Support

But where advocates of psychoanalysis see a problem about the *communication* of confirmatory data, critics see a problem about the *status as confirmatory* of the data themselves.

Explanation, Evidence, and Bayes's Theorem

We can review these disputes in terms of standard notions from the philosophy of science. Roughly, insofar as we make observations or establish other data, we accept them as they are (this is why we regard them as *data*). Nonetheless, we may want to know *why we might expect these data* to be as we find them. We cannot answer this merely by considering the data on their own; but we can link them with other things by framing *hypotheses* or *theories* that *explain* them, characteristically by representing them as effects of some more encompassing causal mechanisms. In this, we cease to regard the data as isolated phenomena but instead see them as having a particular place in a larger causal pattern, which includes mechanisms that explain why we should expect them to be as we find them.

This means that such explanatory hypotheses or theories are always also *predictive*, in the sense that they represent the data they cover *as to be expected in accord with the working of the mechanisms that explain them*. But for this to be so, the hypothesis (or hypothesized mechanisms) must perforce confer a *probability of the data given the hypothesis* that is higher than the *probability* of *the data given the negation of the hypothesis* (supposing that there is no such mechanism). This is simply what it is for the hypothesis to explain the data in the sense of showing why they should be expected to be as they are. From this, however, it follows that where we advance such a hypothesis in explanation of some open-field data, we should also be able to *confirm* or *disconfirm* that hypothesis—that is, to increase or decrease whatever credibility we initially assign to it. For we can do this simply by updating our assignment of credibility to the hypothesis in accord with Bayes's theorem, as it seems rational to do.

This is because the condition required for the hypothesis to explain the data—that *the probability of the data given the hypothesis* be greater than *the probability of the data given the negation of the hypothesis*—entails that for further data that are as the hypothesis predicts, *the probability of the hypothesis given the data* (the credibility to be assigned to the hypothesis after further data have been collected) will be greater than *the prior probability assigned to the hypothesis*, whatever that was. So even a weakly predictive hypothesis gains credibility—is confirmed—as we observe that things turn out as it predicts. (And the relation is reversed, and the credibility of the hypothesis diminished, where the data turn out not to be as even weakly predicted.) The increase or decrease in probability engendered by further observations, moreover, should be proportional to the probability the hypothesis confers upon them.

On such an account, Freud's hypotheses should be regarded as confirmed insofar as they provide the best available explanation of the data they cover and render expectable; and this is apparently what Freud

was attempting to illustrate by repeated example in publications such as *The Interpretation of Dreams*. His overall claim there was that if an individual seeks to interpret her dreams by pursuing free association in relation to the elements of the dreams, then data will emerge from the free association that allow the dream to be interpreted in the kinds of ways Freud describes. The instances he produced apparently encompass the data explained by his overall account. If so, then, as he claimed, critics ignorant of such data, or the way his hypotheses served to explain them, would not be in a position to evaluate them properly. Criticism of explanatory theories that does not take full account of the observational data upon which they are based is not empirically well-founded.

Critiques by Popper and Grünbaum

Understanding confirmation and disconfirmation in this way also bears on the celebrated critiques of Freud by Karl Popper and Adolph Grünbaum. According to Popper's criterion of *falsifiability*, the only general hypotheses that can be regarded as truly scientific are those that predict data so precisely that they would be conclusively falsified if particular instances of data were not forthcoming. Again, only such precise general hypotheses can be said to explain data, and to be confirmed by them, in a fully scientific way. Such falsifiability should not be confused with the kind of Bayesian confirmation and disconfirmation indicated above. Rather, Popper's account in effect requires that truly scientific hypotheses confer a probability very near to certainty on the data they explain, so that any predictive failure renders the probability of the hypothesis itself negligible. As Popper stressed, the hypotheses of Freud and Darwin mostly do not satisfy this criterion; and for this reason, he regarded them as metaphysical, as opposed to scientific.

This categorization by Popper has very often been used to discredit Darwin or Freud. This, however, misrepresents both Popper's argument and his intentions. Popper not only allowed but emphasized that the general theories he characterized as metaphysical might have true and confirmed instances. His account turned on the difference between singular instances and the unrestricted theoretical generalizations thought of as inductively supported by them. The establishing of singular instances of a general claim (e.g., finding that this or that dream was well understood as a Freudian *wish fulfillment*) could not conclusively verify the unrestricted generalization involved (that all dreams are wish fulfillments). By contrast, establishing a single negative instance (finding a dream—e.g., an anxiety dream or a nightmare—that was not best understood as a Freudian wish fulfillment) could conclusively falsify the generalization.

In light of this, Popper held that genuinely confirmed instances, even in the context of a general theory as scientifically fertile and invaluable as Darwin's, did not suffice to render the generalizations themselves truly scientific. Likewise, and despite criticisms, he stressed that Freud's explanations of instances in *The Interpretation of Dreams* were fundamentally correct and that Freud's description of the unconscious therefore constituted a great discovery. Popper regarded the generalizations of Darwin and Freud as unscientific, not because they lacked confirmed positive instances but rather because the terms in which they were cast were not so precise as to render them conclusively falsified by ostensibly negative ones.

Advocates of psychoanalysis agree with Popper that psychoanalytic hypotheses do not generally yield strict and readily falsifiable predictions like those attainable in the "hard" sciences. This holds, indeed, for the whole of the natural human psychology of desire, belief, perception, emotion, and experience in which we understand one another in everyday life. But in this psychology—and contrary to Popper's claims—our ability to understand one another with precision far outruns our ability to predict one another's actions. This is particularly clear in the case of language, in which we may be able to understand what a person says with as great an exactness as we understand anything in science but without thereby being able to predict what he will say next.

Advocates also stress that in this field examples that admit a degree of generalization may have great importance even if they are subject to restriction. The interpretations of dreams that Popper regards as basically correct already provide grounds for revisions in commonsense psychology that are deep and potentially cumulative and radical. It is highly important—as Popper conceded in speaking of a great discovery—that many, if not all, dreams can be understood in this way. So advocates maintain that if, as Popper agrees, Freud's hypotheses are sufficiently predictive to impose genuine relations of confirmation and disconfirmation on instances

of explanation, and if they provide the best explanations we can frame for the many instances that would be unexpected on other accounts, they should be regarded as constituting a significant and empirically cogent extension of our everyday psychology of motive, by means internal to it.

Comparing Darwin and Freud brings out another methodological point. Where a theory explains data by representing them as produced by an underlying causal mechanism, its predictions entail claims about the mechanism as well. Where the mechanism itself is as yet unexplored, finding out about it becomes part of the project of confirming or disconfirming the theory. Darwin's theory made claims about the mechanisms of heredity—that they produce resemblances between parents and offspring that also allow for modification by natural selection—as did Freud's theory about the neural mechanisms of motivation, dreaming, and mental disorder. The predictive scope of Darwinian theory, and with this the degree of confirmation it enjoyed, steadily increased as the mechanisms of heredity were anticipated and described. Once Darwin's theory could be exhibited as predicting the great fact that the vital and reproductive processes of all living things are regulated by the same basic family of molecular mechanisms, Darwinism attained a degree of confirmation comparable to other paradigms of science. This also illustrated the failure of Popper's criterion, for of course we remain unable precisely to predict the alterations that natural selection will produce (e.g., in the flu virus), even though we may attain detailed understanding of them afterward.

Advocates of psychoanalysis hold that progress in developmental psychology and neuroscience will likewise generate further and significant confirming or disconfirming data. But this will also depend on actively searching for evidence about the working of the mechanisms, as in the Darwinian case of combining physiological research with digging for fossilized instances of forms of life altered or eliminated by natural selection. In psychoanalysis, such digging is shown in looking for the deepest and most significant connections in free associations, memories, motives, and conflicts, as well as in relating these to findings in psychology and neuroscience. In both cases, the digging is mandated by the hypotheses under consideration, which predict the kind of thing (but not in detail) that digging should disclose. And in both cases, digging for evidence is liable to be mistakenly characterized by critics as a search for confirmation in violation of falsificationist strictures.

In contrast with Popper's falsificationism, Adolph Grünbaum provided a critique of Freud on inductivist grounds. Psychoanalytic claims about motives, he argued, were causal claims. Therefore, they required noninterpretive and extraclinical experimental or correlational support, such as the double-blind control group procedures used to test the causal powers of drugs. Again, advocates of psychoanalysis contest these criticisms as failing to take full account of the nature of Freud's hypotheses and data. Certainly, psychoanalytic claims about the role of motives are causal, just as are claims about the roles of perceptions, desires, beliefs, and emotions in everyday life. But then in the indispensible and fundamental psychology of meaning and motive that we naturally use in understanding one another, we already constantly establish such causal connections by intuitive interpretive means. In maintaining that Freud was extending this psychology by means internal to it, advocates of psychoanalysis were also perforce maintaining that other forms of correlational testing, although they might be desirable where possible, were not required.

"Suggestion" as an Alternative Explanatory Hypothesis

In addition, and like many before him, Grünbaum argued that the kind of confirmatory clinical observations to which Freud refers may well be contaminated by the influence of the analyst—for example, by unconscious suggestions with which the patient unwittingly complies. In this, his critique recycled the oldest but also the most influential grounds for rejecting psychoanalytic claims; and similar arguments have also been cited to explain the apparently confirmatory data that sustain competing psychoanalytic schools.

To this familiar charge, advocates reply yet again that it does not take full account of the methodological situation. This is simply that all hypotheses purporting to explain the data in question, including hypotheses that postulate processes such as suggestion, require to be evaluated by how well they actually explain and predict the data involved. In fact, psychoanalytic hypotheses seem to be the only ones that do so with any degree of adequacy. For critics who stress the role of suggestion have provided no account as to how it is supposed to explain psychoanalytic data; and these data, as recorded, show patterns inconsistent with the claim.

Thus, data recorded as explained by early Freudian hypotheses are very often better explained by quite different psychoanalytic hypotheses framed later in order to explain different data. For example, the associations to Freud's dream of Irma's Injection, and to others in *The Interpretation of Dreams*, are clearly better explained by the mechanism of *projection* that Freud was to emphasize in later work than by that of *wish fulfillment* he used in his initial book. (The hypothesis of wish fulfillment does not explain why the derelictions with which Freud is concerned should be assigned to Otto in the dream; that of projection has this as a direct consequence. In addition, projection is the mechanism most explicative of paranoia and so serves to explain the anxiety dreams and nightmares with which critics still tax Freud's first formulation.) Such retrospective but more cogent explanations can be given for a considerably large number of recorded instances of psychoanalytic data, and this pattern is inconsistent with the claim that the data are explained by psychoanalytic hypotheses because they emerge from suggestion in terms of them.

Again, critics often cite the fact that patients of therapists of different schools come to free-associate, and so they produce clinical data using the concepts of those schools. Advocates regard this as showing, as would be expected, that these patients are using the theoretical framework in terms of which they are attaining a deeper understanding of themselves for the purpose for which that framework was devised—that is, to enable them to understand themselves better than they would without such a framework. Likewise, advocates regard the persistence of differing schools of depth psychology as a natural and inevitable consequence of the richness of the data forthcoming in free association (and unconstrained play in the case of the analysis of children) and the motivational complexity of the human mind. As with the long, intricate development of life itself, we may never attain a single best perspective on these things; but that is no reason not to seek such better and deeper explanations as we can frame.

Recent Work in Developmental Psychology and Neuroscience

Advocates and critics of psychoanalysis thus remain in epistemic deadlock about the significance of clinical data. In this situation, it has seemed relevant to both advocates and critics to consider related fields such as developmental psychology and neuroscience, especially since Freud began work as a neuroscientist and framed many of his psychoanalytic concepts so as to be consistent with his own neuroscientific understanding.

Work in both fields has burgeoned in recent decades, and advocates urge that, as was to be expected, the results systematically favor psychoanalytic hypotheses. Thus, attachment theory, developed by the psychoanalyst John Bowlby as an empirical and evolutionary approach to the formation of emotional bonds and to the effects of parental care, has become a well-established field of developmental psychology. It has produced powerful empirical evidence of the influence of parenting on psychological development from as early as four months, and its main modes of testing, such as the "strange-situation" procedure, are in effect ways of determining the role of emotional conflicts that Freud took to be central to mental disorder. So it has arguably produced systematic evidence of the pervasive and lasting role of what Freud described as the earliest parental imagoes in engendering emotional conflict—and with this, evidence for a range of psychoanalytic claims about the importance of such conflict in development.

The observations of attachment theorists also coincide with current hypotheses in developmental neuroscience, which stress how representations of self and other in the infant's relatively unformed cortex are structured under the impact of emotion by interaction with parents and carers over the first year of life. Again, recent work has provided accounts of emotion and dreaming consistent with Freud's and confirmed clinical phenomena that seem to require explanation in psychoanalytic terms. Neuroscientific work seeking to determine the construct validity of psychoanalysis has shown a range of consistencies between key Freudian ideas and recent perspectives on global brain function that have emerged in neuroimaging and Bayesian neuroscience.

The neuroscientific framework of the Bayesian brain, moreover, seems also to provide a simple and direct account of repression and other Freudian concepts, thereby contradicting a range of traditional philosophical arguments against them. As regards the therapeutic effect of psychoanalysis, neuroimaging studies now seek to specify the changes effected by successful psychoanalytic theory in neural systems related to emotion and attachment. So while such empirical work is not directed to the particular clinical observations stressed by Freud, it seems to offer indirect but wide-ranging empirical support

for them, now in terms of the working of the causal mechanisms that underlie them. No doubt philosophical critics of psychoanalysis will respond to these new claims and data; at present, however, they are yet to do so.

Jim Hopkins

See also Abduction and Inference to the Best Explanation; Bayesianism, Recent Uses of; Ego; Explanation, Theories of; Falsifiability; Personal Identity and Trauma; Popper's Philosophy of Science; Pseudoscience; Schizophrenia: Psychoanalytic, Phenomenological, and Contemporary Philosophical Approaches; Scientific Method; Therapy, Psychological and Philosophical Issues; Unconscious

Further Readings

Boag, S. (2012). *Freudian repression, the unconscious, and the dynamics of inhibition*. London, England: Karnac.

Buchheim, A., Viviani, R., Kessler, H., Kachele, H., Cierpka, M., Roths, G., . . . Taubner, S. (2012). Changes in prefrontal-limbic function in major depression after 15 months of long-term psychotherapy. *PLoS ONE, 7*(3), e33745. Retrieved from http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0033745

Chung, M. C., & Feltham, C. (Eds.). (2003). *Psychoanalytic knowledge*. New York, NY: Palgrave Macmillan.

Clark, P., & Wright, C. (1988). *Mind, psychoanalysis, and science*. Oxford, England: Blackwell.

Fotopolou, A., Pfaff, D., & Conway, M. (Eds). (2012). *From the couch to the lab: Psychoanalysis, neuroscience and cognitive psychology in dialogue*. Oxford, England: Oxford University Press.

Freud, S. (1957). Analysis of a specimen dream. In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud: Vol. 5. The interpretation of dreams* (pp. 96–121). London: Hogarth Press. (Original work published 1900)

Grünbaum, A. (1984). *The foundations of psychoanalysis*. Berkeley: University of California Press.

Laplanche, J., & Pointalis, J. B. (1973). *The language of psychoanalysis*. London, England: Karnac.

Lear, J. (2005). *Freud*. New York, NY: Routledge.

Levine, M. (Ed.). (2000). *The analytic Freud*. London, England: Routledge.

Neu, J. (1991). *The Cambridge companion to Freud*. Cambridge, England: Cambridge University Press.

Popper, K. R. (1983). A case of verificationism. In W. W. Bartley (Ed.), *Realism and the aim of science: From the postscript to the logic of scientific discovery* (sec. 18, chap. 2). New York, NY: Routledge.

Solms, M., & Turnbull, O. (2002). *The brain and the inner world: The neuroscience of subjective experience*. London, England: Karnac.

Turnbull, O., & Olds, D. (Eds.). (n.d.) *Neuropsychoanalysis*. Retrieved from http://www.neuropsa.org.uk/journal

Wollheim, R. (1993). *The mind and its depths*. Cambridge, MA: Harvard University Press.

Wollheim, R., & Hopkins, J. (Eds.). (1982). *Philosophical essays on Freud*. Cambridge, England: Cambridge University Press.

Public Goods

Public goods are goods that the free market cannot provide efficiently. Two characteristics distinguish public goods from private goods: *non-rivalry* and *non-excludability*. This entry defines these two characteristics, explains why goods that have these characteristics cannot be efficiently provided by the market, considers the different ways in which the government can provide these goods efficiently, explores the contested and related concept of "merit goods," and considers the relationship between the provision of public goods and the principle of liberal neutrality.

Non-Rivalry

The first characteristic of public goods is non-rivalry. A good is non-rival if one person's consumption of the good does not diminish the benefit that any other person can obtain from consuming the good. An example of a good that is non-rival is a radio station signal. On the other hand, most goods (e.g., clothing) are rival goods. There are some goods that are non-rival up to a point and then become rival when use becomes sufficiently high (e.g., a public park). These are known as *congestible goods*.

The free market will not provide non-rival goods efficiently. When goods are non-rival, it is often the case that the economic benefit from allowing someone to consume another unit of the good is greater than the cost of providing that extra unit (e.g., providing an extra person with cable television access). Therefore, society could obtain greater total economic benefits by having greater production of the good than what the free market would provide.

Non-Excludability

The second characteristic of public goods is non-excludability. A good is non-excludable if it is very difficult or impossible to prevent individuals from using the good once it exists. So, for example, fish stocks in international waters are an example of a non-excludable good. On the other hand, an ice cream cone is excludable since it is easy to prevent someone who did not pay for it from enjoying it.

Different problems of efficiency arise depending on whether the non-excludable good already exists (e.g., natural resources) or whether it still needs to be produced. If the good does not already exist, the inefficiency is straightforward to see. Any private actor who produces the good will not be able to recoup his or her costs. And so, barring altruistic behavior, the private market will not produce this good.

If the non-excludable good already exists and it exhibits rivalry, different efficiency problems arise. First, each person who uses the good will diminish everyone else's benefit from using the good and will not have to pay the costs of doing so. When enough people act in a narrowly self-interested way, game theory predicts that this will lead to the *tragedy of the commons* (an overuse of the resource). Second, assuming that there are barriers to trading the good, it may end up being used by individuals who do not place the highest economic value on it. Finally, individuals may expend effort and resources to appropriate the good early simply in order to prevent others from taking the good.

Types of Economic Goods

We can use these two characteristics to place economic goods in four categories: (1) goods that are excludable and rival (*private goods*), (2) goods that are non-excludable but rival (*common goods*), (3) goods that are excludable but non-rival (*club goods*), and (4) goods that are both non-excludable and non-rival (*pure public goods*).

Note that non-rivalry and non-excludability can *each* provide an efficiency-based justification for the government to intervene in a particular market. This is why some economists classify both club goods and common goods as types of public goods. However, the justification for intervention is often strongest when both characteristics are present (e.g., in the case of pure public goods like national defense).

Table 1 Types of Economic Goods

	Excludable	Nonexcludable
Rival	Private goods Example: ice-cream cone	Common goods Example: limited fish stocks in ocean
Nonrival	Club goods Example: scrambled radio signal	Pure public goods Example: national defense

Source: Author.

Efficient Government Provision

Faced with the private market's inability to provide public goods efficiently, it is not always the case that the government can do better. However, the government has three general techniques it can use to efficiently provide public goods. First, the government can conduct a cost–benefit analysis to measure the willingness to pay of citizens for different types and different quantities of the public good. Second, the government can rely on democratic processes to determine the right level of the public good. Third, different localities can produce different public goods and people can then sort themselves based on the taxes they are willing to pay for the public goods provided.

The success of each of these methods is subject to a wide variety of caveats. Moreover, given imperfect information, often some taxpayers will effectively subsidize others' public good consumption. Finally, the inefficiency of government provision is an important issue. Often economists advocate government *financing* rather than government provision of public goods.

Liberal Neutrality

However, even if the government can efficiently provide a public good, it is not always clear that it should do so. A key concern has to do with the principle of liberal neutrality, which states that the government should not favor one conception of the good life over another. So, for example, even though many religious institutions exhibit non-rivalry, liberal governments do not directly support them. There is also controversy about how a neutral government can justify

building a baseball field rather than a children's recreational area or, even more controversially, whether it can justify publicly funding art.

Merit Goods

Just as there are public goods that the government should arguably not provide, there are also private goods that the government arguably should provide. For example, elementary school education is both excludable and fairly rival. Yet most believe that government should provide this type of good. Some economists refer to private goods that the government should nevertheless provide as *merit goods*. Some even expand the definition of public goods to include merit goods. The justifications for classifying a good as a merit good are controversial and include equity concerns, paternalism, and community values.

Joseph Mazor

See also Common Goods; Cost–Benefit Analysis; Homo Economicus; Markets and Economic Theory; Pareto Optimality; Social Choice Theory

Further Readings

Batina, R. G., & Ihori, T. (2005). *Public goods: Theories and evidence*. Berlin, Germany: Springer-Verlag.

Brighouse, H. (1995). Neutrality, publicity, and state funding of the arts. *Philosophy & Public Affairs, 24*(1), 35–63.

Dworkin, R. (1985). *A matter of principle*. Cambridge, MA: Harvard University Press.

Mankiw, G. N. (2011). *Principles of economics*. Mason, OH: South-Western.

Van Eecke, W. (2006). *An anthology regarding merit goods: The unfinished revolution in economic theory*. West Lafayette, IN: Purdue University Press.

Public Reason and Justification

This entry introduces the idea of public reason and justification that plays a central role in contemporary liberal theories of political legitimacy. The idea of public reason is important for the philosophy of the social sciences since it delineates an important element of public, or collectively constituted, life.

The liberal idea of public reason has its roots in Immanuel Kant's 1784 essay "What Is Enlightenment?" In Kant's usage, private reason is addressed to the members of a particular association, such as a state or church, and presupposes the existence of an authority that is recognized by the members of that association, such as a set of positive laws or scriptures. Public reason, in contrast, acknowledges no other authority than that of reason itself and is therefore addressed, at least in principle, to the entire community of reasoning creatures. Kant holds that the free and public use of reason is a necessary condition for the progress of enlightenment.

The liberal idea of public justification is associated with the revival of contractarian thought that followed the publication in 1971 of John Rawls's *A Theory of Justice*, in particular with the emphasis that Rawls places on the "fact of reasonable pluralism" in his *Political Liberalism* (1993). Rawls argues that disagreement about fundamental questions is a permanent feature of life in modern democratic societies and that we should therefore distinguish between public reasons, which are accessible to all citizens, and nonpublic reasons, which are only accessible to those who endorse a particular moral, religious, or metaphysical view. He holds that the exercise of political power is only legitimate if it can be justified through the use of public reasons—that is, in terms that transcend or abstract away from the more "comprehensive" matters on which they disagree. Such a liberalism would be "political, not metaphysical."

The idea of public justification is closely related to the traditional liberal idea of government by consent, and like all theories of consent, it contains a number of ambiguities. The areas of ambiguity are explained in what follows:

a. *Content:* It is not clear to what extent the content of public reason is defined by the beliefs that are *actually* shared by a given community and to what extent it is defined by the beliefs that they *would* share if they were fully rational, fully reasonable, and/or fully informed. Similarly, it is not clear whether the content of public reason varies over time and from place to place and to what extent the use of public reason should be expected to yield unique and

stable conclusions. To the extent that public reason is defined empirically and contextually, it will of course be subject to the various kinds of ignorance and bias that exist in all human communities. Moreover, because few if any substantive beliefs are universally shared, such an approach threatens to reduce the content of public reason to an empty set. Conversely, to the extent that public reason is defined normatively and counterfactually, its content is likely to be determined by exactly the kinds of controversial doctrines that it is supposed to transcend.

b. *Domain:* Rawls and many other political liberals argue that only the basic principles or "constitutional essentials" of a liberal regime—those rules that define the basic rights and duties of citizens—need to be publicly justified and that ordinary matters of legislation need not meet this standard. However, it is not clear why the domain of public reason should be limited in this way or that "ordinary" and "constitutional" questions can be easily distinguished from one another. Some political liberals therefore hold that all decisions in which the use of coercive force is authorized must be publicly justified.

c. *Jurisdiction:* Some political liberals hold that the norm of public justification only applies in official contexts—for example, among public officials or voters—and not in the "background culture" in which open discussion among citizens takes place. Others hold that public reasons only need to be invoked at the moment when a political decision is made and that no constraints should be placed on political debate up to that point. Many critics of the idea of public justification insist that despite these limitations it would have a chilling effect on discourse in the broader public sphere and would discriminate in particular against religious believers.

d. *Motivation:* Many political liberals hold that public justification requires that the reasons that are offered in support of a given set of policies be shared by those who are bound by them. Others argue that it is enough for public reasoning to converge on a set of policies even if different citizens endorse them for different reasons. Relatedly, some political liberals hold that citizens must sincerely endorse and be motivated by the public reasons that they offer, whereas others hold that public reasons only need to be acceptable to those to whom they are offered.

e. *Exceptions:* Rawls argues that it may be permissible to appeal to nonpublic reasons in political debate provided that sufficient public reasons are offered "in due course," a stipulation that he refers to as the *proviso*. The proviso is meant to lessen the motivational strain that some people—paradigmatically, religious believers—might otherwise feel in being asked to adhere to the norms of public justification. Other political liberals insist that the use of nonpublic reasons is never justified in public life and that the appeal to such reasons should therefore be confined to the background culture.

Eric MacGilvray

See also Epistemic Approaches to Democracy; Reflective Equilibrium; Social Contract Theories; Social Norms; Social Rules

Further Readings

Gaus, G. F. (1996). *Justificatory liberalism: An essay on epistemology and political theory.* New York, NY: Oxford University Press.

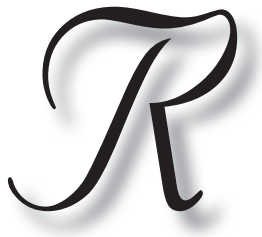
Greenawalt, K. (1995). *Private consciences and public reasons.* New York, NY: Oxford University Press.

Kant, I. (1970). An answer to the question: "What is enlightenment?" In H. S. Reiss (Ed.), *Political writings* (H. B. Nisbet, Trans.; pp. 54–60). New York, NY: Cambridge University Press. (Original work published 1784)

Larmore, C. (2003). Public reason. In S. Freeman (Ed.), *The Cambridge companion to Rawls* (pp. 368–393). New York, NY: Cambridge University Press.

Rawls, J. (1993). *Political liberalism.* New York, NY: Columbia University Press.

Rawls, J. (1999). The idea of public reason revisited. In S. Freeman (Ed.), *Collected papers* (pp. 573–615). Cambridge, MA: Harvard University Press.



Race, Theories of

The modern view of humankind as naturally divided into different "races"—with population groups roughly linked with continentally demarcated origins (white, brown, black, red, yellow)—has had a huge influence not just on popular consciousness but also on academic theory and social structure. For hundreds of years, most Europeans took for granted not merely racial division but also racial hierarchy: Whites were seen as the race superior to all others. Since World War II, biological racism has officially been largely discredited, but controversy continues on how race should be regarded. This entry surveys some of the questions and problems raised for social theory by race as an issue.

Racist Theories Versus Antiracist Theories

Historically, most theories of race have been racist; it is only in recent decades that antiracist theories have become the norm. While all peoples are obviously capable of being racist, it is White racism that has historically had the greatest global impact, because of the shaping of the modern world by European expansionism.

Racist theories vary tremendously in their details, but in general, they make two key assumptions: (1) humankind can be divided into discrete subgroups—"races" and (2) these races are hierarchically positioned with respect to each other—that is, some races are superior to others. The metrics of this putative superiority have usually been cognitive (degrees of intelligence) and characterological (moral vices and virtues). However, aesthetic and physical criteria have also sometimes been employed, with the "superior" race being seen as more beautiful and physically more able than the "inferior" races. In White racism, Whites are placed at the apex of humanity, superior to the "colored" races. But intra-European racism has also demarcated (a) Aryans from Jews, Slavs, and Romani; (b) Nordics from Alpines and Mediterraneans; and (c) Anglo-Saxons from Celts and Latins.

The explanations given in racist theory for this hierarchy have included theological, biological, and cultural accounts, sometimes individually and sometimes in combination. Racist theories that posited separate origins for the different races (polygenesis) were always in tension with Christian orthodoxy, and they became even harder to defend after Charles Darwin's work. Thus, biological ("scientific") racism post-Darwin posited instead radical "racial" variation within a human race conceded to have a single origin (monogenesis). However, though no longer respectable in the academy, polygenetic theory can still be found in the ideas of some contemporary racist fringe groups—for example, the idea of non-Whites as "mud people" not descended from Adam and Eve. Many theorists argue that since World War II, the dominant variety of racism has become cultural racism, where the marker of inferiority is now a deficient and inassimilable culture rather than a deficient body.

Antiracist theories will always reject the second assumption (racial superiority and inferiority); however, they have diverged on the first assumption

(that humankind can be divided into races). Between World War I and World War II, most liberal antiracists would have affirmed the equality of all races while agreeing that humanity was naturally racially divided. But with the discrediting of biological racism by the Holocaust, many antiracists in the scientific community began to develop a more radical position—not only was there no racial hierarchy but there were no biological races to be hierarchically arranged in the first place. The human race was a continuum of varying traits and "clines," with no natural discrete divisions, and what had been thought to be "natural" demarcations were actually arbitrary social cuts, with more genetic variation to be found within groups categorized as "races" than between them. The phrase now standardly used is that race is "socially constructed"—an artifact of particular social systems and social divisions.

Periodization and Origins of Race and Racism

The discrediting of race as biology naturally raises the question of what explains the origins of racial thinking and the genesis of race as a social category. Unfortunately, there is no consensus on this issue. A wide range of theories have been put forward: (a) innate human ethnocentrism mutating into a more dangerous "racial" form, (b) European color symbolism (for White racism specifically), (c) sociobiology, (d) Marxist political economy, (e) psychosexual drives, (f) cognitive psychology, and (g) group conflict theory. Some theorists have been insistent that racism is distinctively modern or, at the earliest, late-medieval, with seemingly contrary evidence from the classical Greco-Roman world to be categorized as color prejudice or ethnocentrism rather than racism. Other scholars have disagreed, contending that racism, albeit not in our contemporary form, goes all the way back to antiquity. Within this latter group, racism has been seen as distinctively Western by some (a Greco-Roman product) and as generally diffused within ancient civilizations (China, India, Japan, the Arab world) by others.

"Race" Today

Contemporary antiracism is divided on whether the social construction of race should be taken to imply that race does not exist at all, and should be eliminated from our vocabulary (eliminativism), or whether races should be seen as nonbiological social entities that are brought into existence by this history of racism and so have a social reality that requires retention of the language of race (anti-eliminativist constructivism). So *race* as a term would now be functioning differently, referring to a socially created group rather than a natural one. Often this division on the metaphysics of race carries over into a corresponding division on public policy recommendations—the question of whether public policy should be color-blind or color conscious. Advocates of the former view would argue that the language of race helps keep racial thinking alive and only perpetuates social division. For the latter, however, racial division objectively exists because of the legacy of White domination, independently of whether it is acknowledged or not, and the language of race is indispensable both for tracking ongoing racial injustice and for devising measures and mobilizing political support to eliminate it.

Charles W. Mills

See also Essentialism; Eugenics, Old and Neoliberal Theories of; Intelligence; Kinds: Natural Kinds Versus Human Kinds; Oppression; Prejudice and Stereotyping; Racial Critiques of Social Science Applications; Social Constructivism

Further Readings

Back, L., & Solomos, J. (Eds.). (2009). *Theories of race and racism: A reader* (2nd Rev. ed.). New York, NY: Routledge.

Boxill, B. (Ed.). (2001). *Race and racism.* New York, NY: Oxford University Press.

Mills, C. (1997). *The racial contract.* Ithaca, NY: Cornell University Press.

Taylor, P. C. (2013). *Race: A philosophical introduction* (2nd Rev. ed.). Malden, MA: Polity Press.

Ward, J. K., & Lott, T. L. (Eds.). (2002). *Philosophers on race: Critical essays.* Malden, MA: Blackwell.

Racial Critiques of Social Science Applications

We often classify each other by race and sometimes treat people of different races differently. Race is often used in the social and biomedical sciences to describe or explain differences between people in socioeconomic status or health. In order to use race as a variable in studies of social or biological

differences in a population, each member must be assigned to one or more of a fixed number of racial categories. This entry describes how these assignments are made and the issues of racial definition and assignment raised by the use of race in the social and biomedical sciences.

Racial Differences

Blacks are reported to have a lower median income and higher rates of morbidity and mortality than Whites in the United States and the United Kingdom. However, the direction and magnitude of the differences depend on how the racial categories are defined and how members of the population are assigned to each racial category. Different assignments to the Black or White category can result in larger or smaller racial differences in socioeconomic status or health. If more members with large incomes count as Black and fewer as White, the racial difference in median income decreases.

Assigning Race

The current official U.S. standard for collecting and reporting data on race lists five categories but is silent on how members of the population should be assigned a race and how to decide whether a member is one race, such as Black, rather than another, such as White. While the U.S. Bureau of Census assigns members a race based on the race they report themselves to be, the National Center for Health Statistics assigns them a race based on the race on their birth or death certificate. However, a member's self-reported race can be different from the race on his or her birth or death certificate. As a result, the member can be assigned one race at birth, a different race as a young adult, and a different one again when he dies.

When measuring racial disparities in income or health, social and biomedical scientists routinely assign members of the population the race they assign themselves. When measuring racial differences in mortality, however, they routinely treat a decedent's race on his death certificate as his apparent race and his self-reported race on an earlier population survey or birth certificate as his actual one if the two are different. They assume that a member's race can't change but can be misassigned or misreported and there can be errors in a count of the number of members of each race in a population.

Error Measurement

Bias or error in the measurement of a variable is only possible if the variable has an actual value from which the apparent value can diverge. If T is the actual value of a variable V, O the observed or apparent value, and e the error in the measurement of the variable, then e is equal to the difference between O and T, but e has no value unless V has an actual one. If beauty were simply in the eye of the beholder, then no measurement of the beauty of members of a population could be biased or in error and no report could over- or undercount the number of members who are beautiful, since there would be no difference between a member's observed and her actual or true beauty. For a measure of the race of a member of the population to be in error, he would have to have an actual race different from his observed or apparent one.

Race cannot be measured like height or body weight, but a member of the U.S. or U.K. population is not whatever race he takes himself to be. There is a difference between being and appearing to be Black or being White and passing as White, but the difference can vary, and a member's actual race is sometimes taken to be his self-reported race and, at other times, his other-reported or parent's race.

Actual Race

Which race is a member's actual race can vary from one socioeconomic or health risk to another. A member's self-reported race can be his actual race in relation to his risk of obesity and his other-reported race in relation to his risk of unemployment. However, since many social scientists rely entirely on self-reported race in their studies of racial differences, their studies of populations in which self-reported and other-reported race are frequently different can be unreliable.

Other-reported race measures differences in risk better than self-reported race if the risk varies with a member's exposure to discrimination, since it is not based on the category the member assigns to himself or herself but on the one he or she is most often assigned by others. As a result, whenever differences between groups in risk are primarily due to racial discrimination, a member's self-reported race should be taken to be his or her actual race only if his or her self- and other-reported race are the same. No one measure describes or explains every risk better than all the others.

We cannot capture something as complex as race with a single way of defining race or by assigning a person to a racial category. Race, like many of the categories social and biomedical scientists use in their studies, such as class and marital status, has a number of dimensions, and one definition of the category can be a better measure of one and a worse measure of another. Allowing the methods of assigning race to vary follows a recommendation of the National Research Council that reports on race should rely on multiple methods of measurement rather than one, since no one measure can capture all the different ways race can affect the health or well-being of members of a population.

Michael Root

See also Genetic Indeterminism of Social Action; Identity, Social; Implicit Bias and Social Cognition; Prejudice and Stereotyping; Race, Theories of

Further Readings

Blank, R., Dabady, M., & Citro, C. (Eds.). (2004). *Measuring racial discrimination*. Washington, DC: Committee on National Statistics, National Research of the National Academies, Panel on Methods for Assessing Discrimination.

Edmonston, B., Goldstein, J., & Tamayo, J. (Eds.). (1996). *Spotlight on heterogeneity: The federal standards for racial and ethnic classification*. Washington, DC: National Academies Press.

Harris, D. R., & Sim, J. J. (2002). Who is multiracial? Assessing the complexity of lived race. *American Sociological Review, 67,* 614–627.

Office of Management and Budget. (1997). *Revisions of the standards for the classification of federal data on race and ethnicity*. Retrieved from www.whitehouse.gov/omb/fedreg/1997standards.html

Office of National Statistics. (n.d.). *Presenting ethnic and national groups data*. Retrieved from www.ons.gov.uk/ons/guide-method/classifications/archived-standard-classifications/ethnic-group-interim-classification-for-2001/presenting-ethnic-and-national-group-data/index.html

Oliver, M., & Shapiro, R. (1995). *Black wealth/White wealth: A new perspective on racial inequality*. New York, NY: Routledge.

Root, M. (2007). Race in social science. In S. P. Turner & M. W. Risjord (Eds.), *Philosophy of anthropology and sociology* (pp. 735–754; Handbook of the Philosophy of Science Series). Amsterdam, Netherlands: Elsevier.

Smedley, B., Stith, A., & Nelson, A. (Eds.). (2003). *Unequal treatment: Confronting racial and ethnic disparities in health care*. Washington, DC: National Academies Press.

Rational Choice and Political Science

Rational choice theory has been used extensively in political science. While it remains a controversial method in the eyes of many, it is a mainstream method that has influenced thinking in political science well beyond its use in formal or game-theoretic articles. This entry first defines rational choice in the context of political science and then examines its influence through Social Choice Theory, in legislative studies, through agency models, in pressure politics, and in constitutional arrangements and the state.

Rational Choice Theory

Rational choice theory is a method of analysis that assumes that agents (people, organizations, groups) behave in a consistent manner, which in turn assumes that they have complete and transitive preference orderings. The consistency requirement of rational choice is simply a predictive tool that enables researchers to examine agent behavior in one context and then, armed with preferences derived from those circumstances, predict behavior in another context. Although agents are not always consistent in this way, when they are, the method enables explanation, and when they are not, it gives political scientists a starting point for a new set of questions. In political science, these preference orderings are virtually always appended to agents in roles and not to biological human beings. Thus, they are applied to "democratic politicians," "presidents," "bureaucrats," and citizens, not to actual people.

Social Choice Mechanisms

The study of voting and elections begins with considering the role of electoral systems. Social Choice Theory is normative, asking how a collection of agents, each of which has well-defined preferences over a set of alternatives, should determine the collective preference. Kenneth Arrow tried to answer the problem by setting a minimally demanding set of normative criteria and then examining which rules satisfied them. His surprising answer is that no decision rule, producing a rational collective ordering from individually rational orderings, satisfies his criteria. This result launched Social Choice Theory. Following from Arrow, it has been found that certain restrictions on the allowable set of individual

preferences will yield rational collective preferences: importantly, the requirement that preferences are single peaked. In that case, majority rule yields a rational group ordering that then reflects the preferences of the median voter. This result assumes that politics can be represented in a single, say, left–right, dimension. The Median-Voter Theorem has subsequently been used in multiple contributions across the whole gamut of political science.

A second set of corollaries from Arrow's theorem shows that all systems of voting are manipulable: That is, voters might vote strategically, ordering their preferences differently than how they really feel, in order to help a preferred candidate (not necessarily their most preferred candidate) to win. A third set suggests that if politics is multidimensional, then with a moderately large set of issues and a moderately large set of voters, we cannot predict a winner. This occurs because any bundle of policies can be beaten by a new bundle. This led to the key question: Why is there so much stability in politics? The answer is given by political institutions.

Legislative Studies

Kenneth Shepsle introduced the idea of *structure-induced equilibrium,* arguing that the committee structure of Congress turned a multidimensional issue space into a single issue space (or at least a smaller number of issue spaces). This allowed the median voter to dominate in committees, ensuring greater stability. Later contributions examined more complex rules for committees and logrolling across committees (which can introduce instability). In other systems, parties and the domination of the executive (cabinet ministers) in different issue domains bring about stability. A large part of legislative studies is now dominated by considerations from this perspective of the structures that lead to government policy.

Agency Models

Agency models examine the relationship between principals and their agents and have been extensively applied to executive–legislative relations and the bureaucracy. Principals use agents to carry out their wishes, but agents might have their own political interests (they might shift policy) and might not always be competent (they might shirk). The basic agency problem is asymmetric information, where principals do not know precisely what agents are doing; a related problem is adverse selection, where those least qualified for positions work hardest to attain them. Agency models examine these problems. They are applied to (a) the selection of candidates for elected office, (b) how legislatures monitor the executive, (c) how elected politicians monitor the bureaucrats, (d) help explain another role for legislative committees, (e) the function of parties, (f) why prime ministers reshuffle their cabinets so often, (g) failures in government policy, and (h) many other features of political life.

Voter–politician models examine not only how voters punish bad politicians but also how bad politicians try to behave like good ones until it is too late to punish them. This has obvious application to term limits, where a state governor or a president can only serve two terms and hence cannot be punished for actions in their second term. It also applies to politicians coming to the end of their careers. More general models of electoral politics examine coalition formation, where constraints include the desire to maximize the amount of power each party in the government enjoys while finding partners with adjacent ideological positioning. Median-voter models are important in these accounts.

Pressure Politics

The behavioral revolution in political science in the 1960s led political scientists to look beyond formal institutional actors to other agents in the political process. Organized pressure groups affect policy formation through the pressure they place on both elected politicians and civil servants. Following from Mancur Olson's *Logic of Collective Action,* rational choice writers have shown that groups will differentially mobilize depending upon the collective action problems they face and the resources they have to overcome those problems. One consequence is that businesses face fewer problems due to their size and concentration, while citizens have diffuse interests. Later models have examined the role of information and the interplay of special interests and citizen demands on vote-maximizing politicians.

Constitutional Arrangements and the State

Rational choice has modeled the most efficient organization of the state in terms of fiscal federalism, where political responsibilities follow fiscal responsibilities. In recent years, agency models have examined the development of the state. The formation of economic organizations such as guilds, kings giving power to parliaments, and the rise of democracy

have been modeled as attempts to overcome commitment problems associated with long-term contracts.

Rational choice is now part of the heartland of modern political science, and its influence extends far beyond articles that can be seen as exclusively formal. The questions its methods have raised are central to the modern enterprise of political science.

Keith Dowding

See also Behavioralism in Political Science; Collective Rationality; Judgment Aggregation and the Discursive Dilemma; Models in Social Science; Preference; Rationality and Social Explanation; Social Choice Theory

Further Readings

Dewan, T., Dowding, K., & Shepsle, K. (Eds.). (2009). *Rational choice politics* (4 vols.). Los Angeles, CA: Sage.

Mueller, D. C. (2003). *Public choice III*. Cambridge, England: Cambridge University Press.

Olson, M. (1965). *The logic of collective action: Public goods and the logic of groups*. Cambridge, MA: Harvard University Press.

Shepsle, K., & Weingast, B. (1981). Structure-induced equilibrium and legislative choice. *Public Choice, 37,* 503–519.

Rational Expectations

This entry deals with rational decision making under risk or uncertainty; it is not about what is known as rational expectations theory or rational expectations hypothesis in economics, developed by John Muth and others, concerning people's expectations of how economic variables will turn out in the future.

The prevailing theory of rational expectations is expected utility theory. According to this theory, rational decision makers choose that option among those available to them that maximizes expected utility—that is, the sum of the probabilistically discounted utilities of its possible outcomes.

The idea was first introduced by the Swiss mathematician and physicist Daniel Bernoulli (1700–1782) to solve what is known as the St. Petersburgh Paradox. A fair coin is to be tossed until it shows heads for the first time. If the first head appears on the kth trial, a player wins $\$2^k$. How much should one pay to participate in this lottery? Its expected *monetary* return is infinite—that is, $(\frac{1}{2})\ \$2 + (\frac{1}{4})\ \$2^2 + (\frac{1}{8})\ \$2^3 + (\frac{1}{16})\ \$2^4 + \ldots + \ldots\ldots\ldots\ldots + \ldots$, —but surely no one would pay all the money they can lay their hands on to play this lottery. What Bernoulli suggested was that one should evaluate the lottery instead in terms of its expected *utility,* where the utility of any given amount of money increases with its amount but at a *diminishing* rate.

Independence Axiom

Such an account of expected utility gets fully axiomatized for the first time in John von Neumann and Oskar Morgenstern's (vNM) 1947 book. The key assumption on the vNM account is what is known as the independence axiom. Although the axiom is buried in a definition of an equivalence class of lotteries, E. Malivaud diagnosed where the axiom was to be found. As it turns out, there are many formulations of the axiom. A relatively transparent version is the following:

> Suppose that a gamble A is indifferent to a gamble B. Then a gamble (A with probability p and C with probability $1 - p$) is indifferent to a gamble (B with probability p and C with probability $1 - p$).

Critique

A deluge of questions were originally posed by researchers regarding the plausibility of the independence axiom, whether the axiom was *descriptive* of how people choose among gambles or normative—that is, how they ought to choose. At a famous conference in Oslo in 1982 (on the Foundations of Utility and Risk Theory), many of the most distinguished decision theorists met (e.g., Amartya Sen, Kenneth Arrow, and Maurice Allais), and they were basically skeptical about the descriptive validity of the axiom, but the French theorist Maurice Allais was also skeptical about the normative status of the axiom.

Mark Machina presented a paper titled "Utility Theory Without the Independence Axiom," which sought to develop an account that avoided the independence axiom. His paper was so impressive that he was asked to come on the final plenary session to defend his view, even though he was a fresh PhD student from MIT. Virtually all took the view that the axiom was suspect as a descriptive axiom but was clearly very well-grounded as a normative axiom. Similar views were implicit in the work of

Daniel Kahneman, Amos Tversky, and many others. The dean of all decision theorists, Leonard Savage, was exposed to the examples that Allais constructed, chose *against* the independence axiom, and then announced that he was grateful for his own theory for setting him right!

However, Daniel Ellsberg wrote a very important paper in which he insisted that the axiom was suspect as a normative requirement. The trouble with the axiom, according to Ellsberg, could be illustrated by the following example: Suppose one has to draw, from each of two urns, a ball, where there are 30 red and 60 black or yellow balls, and then one has to draw again, either from g_3 or from g_4.

	(30)	(60)		
	Red	Black	Yellow	Range of Expectation
g_1	\$100	\$0	\$0	$33^{1/3}$
g_2	\$0	\$100	\$0	$0–66^{2/3}$
g_3	\$100	\$0	\$100	$33^{1/3}–1$
g_4	\$0	\$100	\$100	$66^{2/3}$

That is, the agent is to choose between g_1 and g_2, and again between g_3 and g_4, in each case based on drawing a ball at random from an urn that contains red, black, and yellow balls in the proportion specified: 30 of the balls are red, and 60 are either black or yellow, but how many of each of these two colors is not specified. Many people prefer g_1 to g_2, but they also prefer g_4 to g_3. As suggested, it would seem that they preferred determinate odds to ambiguous odds (as expressed in the last column), or perhaps the explanation is that they rank them in terms of minimum expected return odds to ambiguous odds (as expressed in the last column), or perhaps the explanation is that they rank them in terms of minimum expected return. What is clear is that it can be shown that such a preference ordering violates a version of the independence axiom. Many found this a much more convincing example of a plausible normative violation of the independence axiom.

There were also a number of developments in non-Independence theories as early as the 1950s. Harry Markowitz developed a separate account of how to deal with risk. This work contributed to his receiving the Nobel Prize in Economics. Essentially, he proposed a mean/variance model of evaluation. There was also an important paper on the theory of risk by Alexander Pollatsek and Amos Tversky, which attracted considerably less attention than it should have. The idea was that options, conceived as probability distributions over monetary amounts, could be ordered in terms of risk, expressible as a linear combination of the mean and the variance of the distribution. As they make clear, this cannot easily be squared with expected utility theory. The utility function thereby defined "tips" over and goes into negative territory, which is not permissible.

The discouraging fact, however, is that the serious concerns raised by the Pollatsek and Tversky paper (and other non–expected utility models) have been largely ignored. Virtually all decision theorists simply take the expected utility theory and the independence axiom for granted.

Edward McClennen

See also Allais Paradox; Decision Theory; Homo Economicus; Preference; Probability; Risk; Social Choice Theory

Further Readings

Bernoulli, D. (1738). Exposition of a new theory on the measurement of risk. *Econometrica, 22*(1), 23–36.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics, 75,* 643–669.

Fishburn, P. C. (1982). *The foundations of expected utility: Theory and decision library* (Vol. 31). Dordrecht, Netherlands: D. Reidel.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica, 47,* 262–291.

Machina, M. J. (1982). "Expected utility" theory without the independence axiom. *Econometrica, 50,* 277–323.

Malivaud, E. (1952). Note on von Neumann-Morgenstern's strong independence axiom. *Econometrica, 20*(4), 679.

Markowitz, H. (1959). *Portfolio selection: Efficient diversification of investments.* New Haven, CT: Yale University Press.

McClennen, E. (1990). *Rationality and dynamic choice.* Cambridge, England: Cambridge University Press.

McClennen, E. (2009). The normative status of the independence principle. In P. Anand, P. K. Pattanaik, & C. Puppe (Eds.), *The Oxford handbook of rational and social choice* (pp. 140–155). Oxford, England: Oxford University Press.

Pollatsek, A., & Tversky, A. (1970). A theory of risk. *Journal of Mathematical Psychology, 7,* 540–553.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297–323.

Von Neumann, J., & Morgenstern, O. (1947). *The theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

Rationality and Social Explanation

This entry reviews the role of instrumental rationality in social explanation and the way such rationality expected of social agents fares with respect to different approaches to social sciences (naturalist vs. interpretivist).

Social science has its origin in commonsense views about human behavior (what philosophers call *folk psychology*). Common sense holds, among other things, that people act to get what they want. The notion of *instrumental rationality* is meant to capture this desire-directed pattern of behavior. *Rational* is a term of commendation: To call something rational is to say that it makes sense in some fashion. *Instrumental* rationality is concerned with the fit between desire and action: An instrumentally rational action inherits (some) value from an agent's goals because it (is expected to) advance those ends. Understood this way, instrumental rationality specifies a (family of) relation(s) between each person's ends (and situation) and her behavior. The relation has its roots in a normative claim—people *should* advance their ends—but once identified, it can serve descriptive purposes. Folk psychology suggests that people do act to satisfy their desires and so that actions will (tend to) be instrumentally rational. Rationality, then, is central to the commonsense explanation of human behavior.

Social science should, of course, advance beyond commonsense views. There are, however, different views about the nature of social-scientific explanation. Some see social science as continuous in its goals and methods with natural science (e.g., cognitive psychologists). These naturalists aim to uncover the causal mechanisms behind behavior. Others see social science as discontinuous with natural science (e.g., cultural anthropologists). Humans are distinctive, they note, in that they are self-guiding. Given this sort of agency, the point of social-scientific theories is not to get "beyond" actions to mechanisms but rather to interpret behavior in a way that makes it intelligible.

Most naturalists think that a rationality-based account of behavior can be integrated into the natural-scientific picture of the world. They suppose that the mechanisms that underwrite behavior are characterized (partially and abstractly) by folk psychology: Beliefs, desires, and so on, have real causal powers; such mental states cause behavior just as hitting a piece of glass causes it to break. Causal patterns are often expressed by decision-theoretic formalizations that go beyond common sense, but the basic idea still involves instrumental rationality: Agents do what they expect will satisfy their preferences.

Basing a naturalist account of social science on rationality claims leads to difficulties. Philosophical critics emphasize conceptual problems—for example, worries about the possibility of mental representation and worries that actions and mental states are too conceptually connected for the latter to be causes of the former. Social-scientific critics focus on empirical problems: People don't always fit the rational pattern. In complicated situations, people sometimes become confused and fail to see how actions relate to ends. Even in simple situations, people can fail to attend to some circumstance or end. Several people, for example, have fallen over the edge of the Grand Canyon when a photographer asked them to step back. Both the photographers and their subjects knew they were at the Grand Canyon—that was the point of the photographs—but they were all so focused on the process of taking a good picture that they failed to utilize their locational knowledge. The upshot is that actual human cognition imperfectly tracks the norms of even instrumental rationality.

Naturalists have resources for addressing the foregoing issues. Many philosophers defend the conceptual coherence of rationality-based causal accounts. The basic story is that mental states are realized by entities that can also be investigated at the neurological or physical levels. Fleshed out, this is supposed to ease worries about both mental representation and mental causation. Empirical difficulties do not seem to warrant the wholesale rejection of rationality-based views. People engage in desire-directed action even when they fail to fully live up to rational norms. The unfortunate Grand Canyon visitors were still acting toward one of their ends—getting a good picture. The puzzle is that they focused exclusively on that end (and the beliefs that it made salient). Common sense has rationality doing double duty as both a normative and a descriptive account,

but naturalist can't sustain that intuition: An empirically adequate account of behavior will diverge from its normative roots.

On the interpretive view of social science, instrumental rationality is more than a mere behavioral hypothesis: It is the fundamental standard for intelligibility. Social science begins, then, with a proprietary scheme for making sense of human action. Familiar behavior is already intelligible, so the real work is to assimilate unfamiliar actions (e.g., those in other cultures) to the rational pattern. Interpretation, on this view, is a kind of translation project: Action is a sort of "text" that others attempt to "read." The conceptual background of action is like the language in which an utterance is expressed: Both have conventions and norms. Fluency, in either case, is the capacity to follow (and exploit) those conventions and norms. A social scientist must immerse herself empathetically in the concerns and situations of others in order to understand their "languages" and so their "statements." While this is very different from the naturalist approach, it depends even more strongly on instrumental rationality.

Interpretive social science faces an issue connected with the empirical challenges to naturalist social science: Some actual (sets of) behaviors just don't seem consistent with instrumental rationality. To the extent that someone's range of actions can't be seen as advancing her system of ends, her behavior can't be explained. Consider, again, a person who falls in the Grand Canyon after following the suggestion to "scoot back." Her desire to take a good picture renders her behavior only partly intelligible. Given what she already knows, a backward step thwarts virtually all of her ends (even her photographic goal). If some actions fail to meet the standard of instrumental rationality, then interpretivists will be forced to conclude that some behavior is at least "inarticulate" and probably just unintelligible.

Stephen Ellis

See also Causes Versus Reasons in Action Explanation; Common Sense (in the Social Sciences); Explanation Versus Understanding; Folk Psychology; Goal-Directedness; Naturalism in Social Science; Rational Expectations

Further Readings

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy, 60,* 685–699.

Davidson, D. (2001). How is weakness of the will possible? In D. Davidson (Ed.), *Essays on actions and events* (2nd ed., pp. 21–42). Oxford, England: Clarendon.

Ellis, S. (2008). The varieties of instrumental rationality. *Southern Journal of Philosophy, 46,* 199–220.

Elster, J. (2007). *Explaining social behavior: More nuts and bolts for the social sciences.* New York, NY: Cambridge University Press.

Rosenberg, A. (2007). *Philosophy of social science* (3rd ed.). Boulder, CO: Westview.

Realism and Anti-Realism in the Social Sciences

The problems of realism and anti-realism are diffuse, ontological, and epistemological; roughly, they are problems of determining what exists and what can be known to exist. We are, accordingly, forced into some philosophical minefields with some huge consequences regarding the methodology and the goals and limits of social science. After explaining what common sense understands by realism, this entry reviews the rise of scientific realism and anti-realism; discusses the main issues involved in the realism/anti-realism debate in social-scientific matters—that is, the issues of explanation, objectivity, and what it means to hold that reality is socially constructed; and ends by presenting a third kind of realism, called *critical realism.*

Commonsense Realism

We begin with the realism of the ordinary person, called by philosophers *commonsense realism, naive realism,* or *pre-critical realism.* It is a realism that takes for granted the existence of the world just as it is experienced. Physical "things," rocks, chairs, and gorillas, are "out there," "observable," and all *certainly* exist. This "world" becomes the paradigm for existence claims. But this "world" might also contain ghosts, souls, or God. And this is bound to stimulate further inquiry. All inquiry, from the mundane to the philosophical and scientific, begins with commonsense realism. But philosophical inquiry has had its effects on our understanding of science, perhaps, especially the social sciences.

Indeed, once René Descartes convinced people that they could doubt the existence of nearly everything—the external world, even God—modern

epistemology was born. Speaking broadly, two programs emerged. One followed Immanuel Kant, who held that *scientific* knowledge was restricted to the phenomenal world—the world of sensory experience. Moreover, this world was concept dependent: We experience rocks because we have the concept "rock." As a "critical" philosophy, the *noumenal world,* the world *not* perceivable by mortals (i.e., the world of "things-in-themselves") exists, but it is unknowable. It is the realm of faith, of God, freedom, and immortality. Commonsense realism was rescued for science without compromising either faith in a supreme being or the new deterministic science.

But another philosophy aimed at understanding modern science became historically dominant. Clearly articulated in 1830 by August Comte, it was developed by philosopher-physicists in the last quarter of the 19th century. As *the* philosophy of science, it became hegemonic in the 1930s with the *logical positivism* of the Vienna Circle.

The Rise of Scientific Realism/Anti-Realism

Science had to expunge all metaphysics. There were two main problems. First, naive realism had to be cleaned up. Second, what about "quarks" and the other nonobservables of science? The biggest cleanup was ridding science of commonsense causality. By the time of Comte, David Hume's analysis had succeeded in this. For Hume, causes would no longer be unwitnessable powers—not the powers of Gods, the telos of Aristotle, the "dormative powers" of opium, or, remarkably, the very commonsensible power of, say, a tidal wave or unobservable mechanism. For science, "*A* causes *B*" would not mean "*A* produces *B*." Rendered into the extentionalist logic of *Principia Mathematica,* it would simply mean "If *A,* then *B*." Ernst Mach and Pierre Duhem had already grasped the consequence: Since to explain was "to strip reality of the appearances covering it like a veil," *science did not explain.* Bertrand Russell took a different and retrospectively preferable route: The very *idea* of cause should be expunged from science. A phenomenalism would replace both naive realism and a critical realism that affirmed the knowability of the independently existing world.

In response to Mach and Duhem and with the tools of modern logic, the deductive-nomological (D-N) model of explanation became the normative standard of scientific explanation. Briefly, some event (or law) *E* is explained not by discovering *what brought it about* but by *deducing* it from a set of experimental laws and conditions. To ensure the absence of metaphysics, a distinction was drawn between theoretical terms (*T* terms) and terms in the observation language (*O* language). *T* terms could be part of a theory only if they were given clear empirical reference in the *O* language.

Scientific realism and anti-realism enter when we ask whether *T* terms refer to an existing entity. Realists affirm that they do: The *T* term represents an existing entity. Instrumentalists ("pragmatists") deny this. For realists who accept the D-N model, statements about the behavior of postulated nonobservables are either true or false. Presumably, then, the theory explains by deriving true consequences from true premises. For instrumentalists, a theory is neither true nor false but "adequate" in the sense that it gives "good predictions."

By the 1950s, positivists saw a serious problem with the instrumentalist account of theory. Carl Hempel framed it as a dilemma. If the *T* terms of a theory achieve their purpose—to establish definite connections between observables—they are not necessary. But if they do not achieve their purpose, they are surely unnecessary.

Most philosophers have opted for a realist interpretation of *T* terms. Milton Friedman, the economist, is a famous exception. In a very influential article, he insisted that it did not matter whether the premises of a theory were true or false. All that mattered were "good predictions." This opinion is widely held by social scientists and occurs in textbooks in the social sciences.

Objectivism

In the effort to model social science on natural science, many social scientists followed Émile Durkheim's dictum: Treat "social facts" as "things." Whatever Durkheim may have intended, the analogy to "natural facts" allowed sociology to be a genuine science. *Social facts had objective existence and could be known objectively through observation, reasoning, and logic.* This assumed the *O* language of naive realism and made irrelevant metaphysics, including post-Kantian critical realisms (see below). The judicious use of tests and "operational definitions" ensured empirical reference. If there were problems (and there were plenty of them), these could be left to the philosophers.

Quantitative work provides the best example. It identifies factors in ordinary experience as *variables* to be measured and correlated—"data." Thus, for example, the *dependent variable* might be "mental development," measured by four tests, and the *independent variable* "ethnicity." The regression yields an "explained variance" (R^2) of 11%. It is, however, not clear what this means.

Anti-Objectivism

There is, however, a body of work that takes a different stance toward the commonsense social world. Most critically, it rejects the assumption that this world is "given" and, following Alfred Schütz, holds that "objectivist" sociology ignores its primary requirement: It asks, "What does the social world mean for the social scientist" without first asking "What does the social *mean* for observed actors within this world?" and "What does the actor *mean* in acting within it?" With these questions, the social world is no longer naively accepted just as it appears but demands inquiry into the mechanisms of the activity by which actors understand one another and themselves. It demands asking not just what they believe but the genesis of meaning and the processes involved in reproducing and changing belief and meaning—including here the practices of social scientists. *Verstehen* ("understanding") and the assumption of intersubectivity is methodologically critical.

There are a large number of otherwise very different versions of this view. Symbolic interactionism, ethnomethodology, and the work of Erving Goffman stand out. But the main idea is affirmed also by so-called postmodernists, some Marxists, and the more recent efforts of Pierre Bourdieu, Anthony Giddens, and critical realism. Both groups of writers are vigorously antipositivist. But in challenging objectivism, questions regarding relativism and the reality of the social world were challenged.

The Social Construction of Reality

Commonsense realism left no doubt as to the existence of rocks or trees. Consider then, mortgages, money, husbands, criminals, and a host of other commonsense "kinds." Surely these also exist and are real. But there is a huge difference in the ontology of the "objects" in the two lists. The objects in the first list exist independently of us in the sense that were there no longer persons in the world, these objects would still exist. *This is not true of the other list.* They are "concept and activity dependent." Without us, the paper of the dollar bill would still exist. It would still burn, but it is only money if people accept it in exchange. Indeed, *all our interactions* take for granted a host of interconnected socially constructed institutions and practices that, taken together, constitute our social world and its history. All of these have histories, which as Karl Marx noted, we make, though only with materials at hand.

It was perhaps inevitable, however, that social construction was particularly aimed at practices that were not only taken for granted but were deemed inevitable, bad, and, accordingly, demanded change. This has ignored the fact that no element of the social world escapes social construction in the foregoing minimum sense. While the focus on objectionable constructions has been liberating, there are problems. One is the tendency to ignore the *reality* of the social world, to ignore the *material conditions* of action, and to suppose a subjectivism in which reality is entirely in the mind of actors. But while social construction rejects "objectivism," it requires "intersubjectivity," *not* "subjectivity." The tendency toward subjectivism has led some writers to suggest a radical relativism and to hold that social change is merely a matter of showing that if the object is a social construction it is instantly malleable.

A bigger problem, perhaps, is the collapse of the distinction between "social reality" and "reality"—the view that *everything* is a social construction. Ontologically, this is a form of *philosophical idealism* (as in Bishop Berkeley: "To be is to be perceived") and, more recently, a *linguistic idealism* (in which only what is discursive exists). Idealism in this sense is a contrasting term to realism, that there is a knowable independently existing external world.

It is easy to see what motivates relativism and idealism. The realist supposes that there are true representations of reality. But the realist acknowledges that representations (ideas, concepts, theories, etc.) are socially produced, that we can know an object only under some description, and that we cannot achieve a "God's eye view of the world." Hence, "correspondence" cannot be the *test* of truth. Hence, as there are competing incommensurable descriptions, there are competing socially constructed "worlds."

To be sure, science is fallible; but if, as the pragmatist Charles Sanders Peirce insisted, belief needs to be constrained by "the real" and if competing worlds

("paradigms") are not incommensurable, then they can be evaluated. Indeed, the "quest for certainty" would seem to be an unspoken assumption of views that deny all utility to evidence and argument. Second, we can distinguish between the social construction of a *representation* of reality and the construction of reality itself. Thus, constructing "kinds" (gorillas, criminals, or subatomic particles) may well be open to various alternative conceptual schemes ("theories"), and this is important but, as noted, not fatal. The reality of the putative extension of a kind, like the choice of conceptual scheme, can be determined by evidence and argument. Thus, we have the concept ("idea") of witches, corpuscles, or (appropriating Ian Hacking's example) "satanic ritual child sexual abusers." For the realist, there are no witches, corpuscles, or satanic ritual child sexual abusers. There are, however, gorillas, child abusers, and quarks. And all three existed before we had the pertinent concept.

Here, another contrasting pair is pertinent: the dichotomy between *realism* and *nominalism*. In one of its senses, the realist holds that "the world" (natural and/or social) is structured. "A good account" represents this structure. Nominalists deny this. For them, there is more than one way to provide a good account. The difference in the ontologies of the natural and social worlds suggests that one might be a realist—even a critical realist—as regards one, the natural, and a nominalist as regards the other, the social. That is, perhaps there are "natural kinds"—NaCl, mammals, or oranges—but only "ideal types" as regard social "kinds." Max Weber here comes to mind.

Critical Realism

In the 1970s, *critical realism* began to attract attention. To distinguish it from those realisms that accept the D-N model, we refer to what follows as "critical realism," although this term has both a wider and a narrower reference, including, for example, Roy Wood Sellars, Bernard Lonergan, Roy Bhaskar's wing of the new realist movement, and, indeed, several important writers, such as Giddens and Bourdieu, who seem to share in a consensus on key critical realist themes but have not identified themselves as realist, critical, or otherwise. (Critical realism is also a recent position in economics advocated by Tony Lawson—see the entry "Critical Realism in Economics.")

One obvious point of consensus is the effort to articulate a third way between the neopositivist ("empiricist") philosophy of social science and antinaturalisms, which reject the idea of a human "science." As part of this, it holds to a post-Kuhnian epistemology, which denies a theory-neutral "given" and a God's eye view of the world. Instead, an ontological realism is joined to the hermeneutic tradition in epistemology, providing what might be called a "situated objectivity" and a "modest relativism."

Second, and perhaps most important, is the rejection of the empiricist Humean's conception of cause, and the "event" ontology on which it rests. The main idea is this: Instead of asserting that causality reduces to a regularity between two "observed" events *A* and *B,* the critical realist, following Rom Harré, seeks the *causal mechanism or mechanisms—observed or unobserved*—that *explain* the regularity. The D-N model is firmly rejected. Thus, NaCl dissolves in H_2O *because* of their respective molecular structures. Indeed, for the critical realist, providing an understanding of the contingent regularities found in experience is precisely the goal of theory. Accordingly, critical realists reject the idea that the goal of science is prediction and control.

The foregoing has been meant to apply to social science as well as natural science. But this has been controversial. The socially constructed character of society contrasts sharply with the mind independence of the natural world. This difference has many consequences, ontological and methodological. Rocks and money are both real. But first, the difference between them requires an ethnographic moment for social science in which we investigate the beliefs of members. This is not true of inquiry in natural science. Second, because practices are concept and activity dependent, there is an emancipatory moment in social science: When we act differently, structures change. But third, if independently existing "nature" is the object of inquiry in natural sciences, what is it in the social sciences? "Persons?" "Their actions?" "Social structures?" "Society?" In particular, to what may causal powers be attributed?

Among realists, this question has been contentious. But because critical realism is very much agent oriented (in contrast to "objectivist" sociology), there is a general consensus on two absolutely critical points. First, persons have causal powers: They can make things happen. Second, and as important,

what they do is *not* "determined" in the sense that they could not have done otherwise. This conclusion follows from rejection of the determinist ontology of positivism. There are no "closed systems," and all "events"—including the acts of persons—are outcomes of complex causality understood in realist terms. In what they do, actors are enabled and constrained. But how are they enabled and constrained?

Unfortunately, on this issue, the water is muddy. Perhaps four positions may be discerned.

1. Social structure is real in that it is causally efficacious (e.g., as evinced by Bhaskar). This is essentially the Durkheimian position, *except that* structures (and society) do not exist independently of human activity, the error of reification—treating what is unreal as real. Unlike the "dialectical model" developed by Peter Berger and Thomas Luckmann, action does not *create* structure: For any individual, it is already made. Hence, as medium and product, action can only reproduce or transform it. Some critics have found this position incoherent.

2. At the other extreme (e.g., in Harré), social reality is exhausted by what people do. The roles, rules, and conventions that are implicated in action are not "powerful particulars" but are only "discursive"; hence, they are the "wrong kind of thing" to have causal powers. For Harré, Bhaskar and most critical realists are guilty of reification. According to most of them, Harré lapses into a sociologically impoverished methodological individualism that even disclaims the theoretical use of, for example, class. Critics argue that discursive acts have both concrete meaning and material conditions and consequences. Thus, even if discursive acts sustain, for example, class, assembly-line activities presuppose workers who need a job, and there are material consequences, for example, the autos that are sold for profit.

3. A compromise is suggested by Giddens: Social structure has but "virtual reality," incarnate in the practices and activities of persons. On this view, social structure lacks "substantive" reality and is an abstraction from concrete activity. Since rules and conventions define resources that enable and constrain action and these are only intersubjectively sustained—to avoid reification—this is read by critics as an idealism.

At issue here is the question of whether Harre's realist causal theory of "productive powers" offers a sufficient analogy between causation in the social and natural sciences. The idea of a social mechanism, reproduced by acts of agents, with consequences, generally unintended, is pertinent here.

4. Finally, is a causal theory of explanation the preferred model of explanation? Here, commonsense realism is on the side of the critical realists, not merely in affirming a productive power account of causality but also in making the agency of persons the paradigm of causality. We are not likely to get from social science better explanations of action than common sense gives us. Accordingly, the goal of social science is not the explanation of behavior. Rather, it is, as in natural science, an understanding of the patterns and regularities found in social reality.

Peter T. Manicas

See also Causation in the Social Sciences; Critical Realism in Economics; Explanation, Theories of; Holism, in the Social Sciences; Kinds: Natural Kinds Versus Human Kinds; Logical Positivism/Logical Empiricism; Normativism Versus Realism; Objectivity; Reductionism in the Social Sciences; Social Construction of Reality; Social Ontology, Recent Theories of

Further Readings

Berger, P. L., & Luckmann, T. (1967). *The social construction of reality.* New York, NY: Anchor Books.

Bhaskar, R. (1979). *The possibility of naturalism.* Atlantic Highlands: Humanities Press.

Giddens, A. (1984). *The constitution of society.* Berkeley: University of California Press.

Hacking, I. (1999). *The social construction of what?* Cambridge, MA: Harvard University Press.

Manicas, P. T. (2006). *A realist philosophy of social science.* Cambridge, England: Cambridge University Press.

Schütz, A. (1970). *On phenomenology and social relations.* Chicago, IL: University of Chicago Press.

Searle, J. R. (1995). *The construction of social reality.* New York, NY: Free Press.

Varela, C., & Harré, R. (1996). Conflicting varieties of realism: Causal powers and the problem of social structure. *Journal for the Theory of Social Behavior, 26*(3), 313–325.

Reduction and the Unity of Science

Reduction replaces more with less. In philosophy and science, this means one of three things. In *ontological* reduction, one entity has its place taken by the other (often its own constituents). The other types are *epistemological:* In a *methodological* reduction, one way of studying something is used in favor of another. In a *theoretical* reduction, the work of one scientific theory is done by another.

Each of these suggests a way of *unifying science*. If everything is constituted from a set of basic entities, then we have general ontological reductionism. If there is a single correct method for all science, then we have general methodological reductionism. If all scientific theories reduce to one, then we have general theoretical reductionism. These reductions can, but do not need to, coincide. Even if overall reduction is blocked, partial reductions may be possible.

The history of social science includes various programs animated by ideals of unity or reduction. Auguste Comte wrote of a "social physics" (later called *sociology*) methodologically inspired by natural science, but he saw sociology as being less abstract than physics or mathematics. Karl Marx envisaged a science of history that could achieve the predictive successes of some physical sciences.

This entry looks at early views about reduction and unity and describes the main types of reduction and the major sources of opposition to reduction.

Early Ideas

Notions of unity are very old. Pre-Socratic philosophers proposed that everything was a single substance (water for Thales), a single principle (change for Heraclitus), or something more abstract (ideal forms for Plato, numbers for Pythagoras). That the world was unified made unified knowledge possible. In between these ancient thinkers and the modern Scientific Revolution of the 17th century, various mystical visions of unity were developed, some inspired by the idea that all knowledge was expressible in a single language or code.

Since the beginning of the Scientific Revolution, more modern visions of unity have been articulated and views rejecting unity developed. For Descartes, Galileo, and other "mechanist" thinkers, mathematics provided an inspiration, and they held that systematic knowledge was essentially mathematical. They were motivated by the power mathematical approaches showed in overthrowing ancient views about the universe and in explaining various mechanical and optical phenomena.

Most mechanists thought that the important properties of the physical world were quantitative and geometric ones like position, shape, and velocity and that all material processes could be explained by reference to them. They hoped that *all* physical knowledge could be found by the same means as had worked in areas like astronomy, ballistics, and optics, and so could be unified. Their views combined elements of ontological and epistemological reduction, with the important qualification that most of them were dualists, who thought that immaterial souls (and God) also existed and could not be studied by the same methods as physical things.

Mechanism failed for various reasons, both within physical science and outside it. Isaac Newton's account of motion was a major scientific success but was not strictly mechanist, partly because it referred to forces. His work suggested a different kind of method and also the possibility of other kinds of force. Charles-Augustin Coulomb's later discovery of a law and force of electrostatic attraction, along with increasingly sophisticated biology and chemistry, suggested that the material world was more active, and more varied, than the early mechanists had imagined. This suggested new ways of conceiving unity and different forms of and challenges to reduction.

Ontological Reduction

The most common form of ontological reductionism in contemporary philosophy is *physicalism,* or the view that everything real is in some sense physical. Physicalism is not mechanism, because what is physical includes things not recognized by mechanists (e.g., fields and forces), and it is up to physics, not philosophy, to determine what is physical. Many physicalists support the autonomy of some nonphysical sciences because they think that theoretical reductions are blocked. The success of ontological physicalism is partly attributable to the demise—within science—of metaphysical dualist views, which held that that people were a combination of mental and physical substances. Growing understanding of the brain helped defeat dualism, as did the failure of

vitalism (the view that nonphysical forces operated in some biological processes). Inspired by Newton's example, biologists and chemists had posited additional forces to explain biological and chemical phenomena that resisted mechanical explanation. Later experiments, especially concerning the conservation of energy, largely ruled out vital forces, and advances in quantum mechanics indicated that the structure of the periodic table was a fully physical phenomenon.

Ontological physicalism is dominant, but among the majority who support it, there is debate about whether any epistemological reduction of nonphysical knowledge is possible.

Methodological Reduction

The most successful and widely supported methodological reduction is mathematics, where it is commonly held that all mathematical knowledge can, or at least should, be expressed in axiomatic set theory. Matters are more complicated where theories have empirical content, as in science.

A historically important methodological reduction in philosophy was logical empiricism, which held that claims about material objects could be reduced to claims about the content of sense perception by human observers. Despite some important achievements, this program was defeated and is no longer a serious philosophical option.

The most famous and controversial methodological reduction in the human sciences was behaviorism. Behaviorists held that introspection was scientifically useless, so psychological phenomena should be studied by reference to evidence about publicly observable evidence—primarily inputs (stimuli) and outputs (responses). Behaviorists made important discoveries about motivation and patterns of behavior. But behaviorism has been displaced from its central role by research programs that study states and processes intermediate between input and output. Part of the explanation for this is the rise of methodological approaches that worked out how to infer aspects of the computational properties of intermediate processes from patterns in input and output.

In social science, Comte, John Stuart Mill, and others held that social phenomena were no more than the sum of activity by individual people and should be studied as such. A related view in biology held that living things could be understood in terms of the combined activity of their individual specialized organs. Neither view remained dominant. In psychology, Lev Vygotsky and others challenged it, arguing that higher mental functions depended partly on social processes, including feedback. The notion of feedback was also important for the rejection of simple organ decomposition in biology by cyberneticists. It is now clear that some but not all social phenomena can be understood individualistically.

There is no committed defense of global methodological reductionism in empirical science today. Besides the specific failure of logical empiricism, the reasons for this include the fact that few now think that there is any interesting sense in which there is a distinctively scientific method. Instead, we see occasional and specific reductions between this or that science. Most contemporary debate over reductionism concerns theoretical reductions.

Theoretical Reduction

In the natural sciences, reducing one theory to another usually means showing that under some conditions the one is a special case of the other. For example, Galileo's view that the path of a projectile is a parabola is a special case of Newton's that it is (a section of) an ellipse, under false assumptions such as that gravity operates in parallel lines. Even though the required assumptions are false, the reduction helps clarify how the theories relate. This kind of reduction, where the theories are about the same things, is not usually controversial. Most debate and discussion about reduction in the social sciences has focused on the issue of reduction between theories about different things.

Theoretical reductionism differs from methodological reduction in that it is not a view about how enquiry should be conducted but about how what is discovered can be related. Two sciences may proceed by different methods, but the theory produced by the one might still be reduced to the theory produced by the other. The two main forms of reduction in this intertheoretic sense are "microreduction" and bridging reduction.

Microreduction

In 1958, Paul Oppenheim and Hilary Putnam argued that the history of science showed a trend toward the unification of science by microreduction. One branch of science microreduces to another if the

parts of things described by the first kind are studied by the second and the second can explain any observations that the first can explain. They divided the world into six levels, from "social groups" to "elementary particles," and argued that there was a trend toward unification by microreduction.

This approach to theoretical reduction was partly motivated by a form of ontological reductionism where complex things had parts and there was a most basic physical kind of part, from which everything was made. But it is primarily a view about the relationships between theories.

General microreductionism failed spectacularly. The evidence in favor of it, especially for the social and psychological levels, was never strong, and there is overwhelming evidence that relational and external properties between parts are often scientifically crucial. Various programs of "cybernetic" or "systemic" thinking about biological and social systems have emphasized this and accumulated much of the evidence. Microreduction is unsuitable for describing important "within-level" reductions such as the reduction of Galileo's account of projectile motion to Newton's account of motion, because such reductions do not involve part–whole relationships. Finally, the other main form of theoretical reduction has led to more intense and fruitful debate.

Bridging Reduction

Most philosophical discussion of reduction has been organized around a model of reduction articulated by Ernst Nagel, who held that reduction is the (deductive) *explanation* of a theory by another theory. Nagel distinguishes between *homogeneous* reductions (where the theories share "descriptive terms") and *heterogeneous* ones (where the vocabularies of the two theories differ). In these cases, Nagel proposed that reduction would require additional assumptions in order to connect the theories.

He described an example where subject to a number of assumptions temperature could be related to mean kinetic energy of molecules and the Boyle-Charles law for ideal gases could be reduced to statistical mechanics. Nagel was clear that this reduction did not do away with thermodynamics (the study of temperature) and that achieving the reduction depended on various pragmatic factors, including the state of development of the two theories and the independent plausibility of the additional assumptions. Nagelian reductions can also be *partial*—so different assumptions may be needed in order to deal with nonideal gases, or no reduction may (yet) be possible. Much later discussion of reduction, especially in the philosophy of mind, and about the "special sciences" has ignored these pragmatic and contextual factors in two ways. First, it has focused on idealized or "toy" theories rather than specific actual science, and second, it has treated Nagel's requirement of connectability in narrowly logical terms, as requiring either laws linking the two theories or identities between the kinds referred to. This logical version of the requirement is the main target of the most famous and widely discussed arguments against reductionism. Under this idealized reduction, the reduced theory would become redundant, since *all* of its content could be generated from the reducing science plus the "bridge laws."

What this means is that it is important to be clear whether a specific reduction between actual theories is being discussed or general reductionism possibly about idealized theories, and in both cases, it is necessary to be clear about what counts as a reduction.

Contemporary Opposition to Reductionism

The most important positions opposed to the unity of science, or to reductionism, are various forms of dualism, emergence, and functionalism. These categories are not, as we will see, strictly exhaustive, and so some positions may combine elements of both.

Dualism

Metaphysical, or ontological, dualists hold that there are two different fundamental kinds of substance or stuff. Most commonly these are something material, or physical stuff, and something mental, or immaterial stuff; hence, "mind–body dualism" described what was the case for entities that consisted of both kinds of stuff—such as humans. Instead of two kinds of "substance," some more recent dualists prefer to speak in terms of one and the same individual entity having both physical and mental "properties." Whether this way of putting things amounts to an ontological dualism depends on the applicable theory of properties. From outside of philosophy, the difference is rarely easy to discern. Whether they are ontological or not, property dualisms (or pluralisms) are relatively common. And to

the extent that they are correct, they may present challenges to methodological reduction and theoretical reduction. If minds are not the same kinds of entity as bodies or if mental properties aren't the same kinds of property as physical ones, then it is not obvious that they should be studied in the same way or that theories about them could be integrated.

There are almost no contemporary metaphysical dualists within science. But some other dualisms continue to command respect. One used to be called the distinction between "spirit" and "nature." In the late 19th century, Wilhelm Dilthey argued that the human sciences (*Geisteswissenschaften*) should be contrasted with the natural sciences (*Naturwissenschaften*) because the former are distinctively concerned with interpretation, context, and normativity and because the importance of history and interpretation places limits on how general explanations in the human sciences can be compared with those in the natural sciences. Many still share this vision of a deep and fundamental division between natural and human sciences. More contemporary members of this family of views hold that intentionality or values are not reducible and hence that beings that have either (especially humans) should be studied accordingly.

Recent science has complicated matters for these views. Some theoretical approaches are used on both sides of the imagined division. Evolutionary game theory, for example, unifies aspects of microeconomics, biology, psychology, and political science, among others. The current range of applications of this theory wreaks havoc upon attempts to demarcate the natural and social sciences by reference to theory or methods, because the very same theoretical tools have proved illuminating with respect to phenomena that are human and nonhuman, social and nonsocial, conscious and nonconscious.

Emergence

A property of some system is called *emergent* if it arises out of the interactions of the parts of the system but is not predictable from the properties of the parts. Talk of emergence was first popular in the 19th century and was partly inspired by attempts to make sense of chemistry and biology. Mechanistic explanations of chemical phenomena were unconvincing, and chemical processes appeared to be nonadditive in that some combinations of substances had properties (transparency, flammability) that the individual substances lacked. Various biological phenomena lacked satisfactory mechanical explanations. In both cases, emergent properties, structure, or causal powers (sometimes called "vital forces") seemed to some like attractive hypotheses.

Not all emergentist proposals fared well. A physical explanation for key chemical phenomena was provided by quantum mechanics. Experiments with conservation of energy in living things suggested that there was neither need nor room for vital forces. More generally, what is predictable is partly dependent on the resources and technology available for making predictions. Contemporary computer simulations and models have been used to investigate the results of many kinds of interaction, and the results sometimes are called "emergent." But the existence of the simulation shows that the result is predictable in at least one way. If *emergent* is taken to mean surprising to at least some people, then it is irrelevant to reductionism or unity of science. To count against theoretical reduction, what emerges has to be patterns or regularities on larger scales, perhaps with functional properties.

Functionalism

Functionalism is the view that some states (including mental states) should not be understood by reference to their internal constitution but rather in terms of how they relate to other states and parts of the system they occur in (i.e., in terms of their function). This view suggests that part–whole reduction is irrelevant. Given an additional argument, part–whole reduction is also incorrect. The additional argument is the "multiple-realization" argument, contending that quite different internal constitutions can implement the same function. So, for example, a "memory" might be constituted from neurons or transistors—what makes it a memory is whether it can return stored information appropriately in response to a request from elsewhere in the system. A major motivation for functionalism and the multiple-realization argument was the artificial intelligence (AI) research program, which would have been doomed from the start if psychological states could not possibly occur in systems other than brains.

Functionalism has a more complicated relationship with bridging reduction. The multiple-realization argument is often taken to defeat bridging

reduction but is typically aimed at the caricatured version (idealized toy theories and construing the additional assumptions as laws) described above. Given this picture, some have argued that functionalism produces explanations that are redundant (there is no causal work for the functional states to do because that is all done "at the physical level") or that gratuitously multiply causes (actions are caused both physically and mentally).

Conclusion

Reductions are often contentious. One reason is that they sometimes threaten the autonomy of the science, entity, or phenomenon being reduced. Another is that they can seem to require leaving out or denying the reality of things that some people find important. This partly explains why "reductionist" is often a term of abuse. There is a trade-off here. Genuine reductions are worthwhile intellectual advances. Premature or failed reductions, if taken seriously, might foreclose inquiry where it is needed. That said, debating whether whole categories of phenomena or theory could be reduced without paying attention to real scientific detail has rarely produced anything of scientific interest. Individual cases need to be taken on their specific merits.

David Spurrett

See also Biology and the Social Sciences; Causation in the Social Sciences; Emergence; Individualism, Methodological; Logical Positivism/Logical Empiricism; Microfoundationalism; Mind–Body Relation; Neurath's Unity of Science and the Encyclopedia Project; Reductionism in the Social Sciences; Supervenience

Further Readings

Boyd, R., Gasper, P., & Trout, J. D. (Eds.). (1991). *The philosophy of science: Part 1 Section 3. Reductionism and the unity of science.* Cambridge: MIT Press.

Bunge, M. (Ed.). (1973). *The methodological unity of science.* Dordrecht, Netherlands: D. Reidel.

Carnap, R. (2011). *The unity of science.* London, England: Routledge. (Original work published 1934)

Dupré, J. (1993). *The disorder of things: Metaphysical foundations of the disunity of science.* Cambridge, MA: Harvard University Press.

Fodor, J. (1974). Special sciences, or the disunity of science as a working hypothesis. *Synthese, 28,* 97–115.

Galison, P., & Stump, D. (Eds.). (1996). *The disunity of science: Boundaries, contexts, and power.* Stanford, CA: Stanford University Press.

Nagel, E. (1961). *The structure of science.* New York, NY: Harcourt, Brace & World.

Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Concepts, theories, and the mind-body problem* (pp. 3–36; Minnesota Studies in the Philosophy of Science, Vol. 2). Minneapolis: University of Minnesota Press.

Pombo, O., Torres, J. M., Symons, J., & Rahman, S. (Eds.). (2012). *Special sciences and the unity of science.* Dordrecht, Netherlands: Springer.

Rahman, S., Symons, J., Gabbay, D. M., & van Bendegem, J. P. (Eds.). (2009). *Logic, epistemology, and the unity of science* (Vol. 1). Dordrecht, Netherlands: Springer.

Spurrett, D. (2006). Reductionisms and physicalisms. *South African Journal of Philosophy, 25*(2), 163–174.

Reductionism in the Social Sciences

There are several ways the issue of reductionism in the social sciences could be pursued. A traditional philosophical approach might provide broad philosophical argumentation about the ontology of the social world and our knowledge of it, relying on conceptual considerations. This entry does not pursue such an approach, which embodies ideas that have largely been rejected in philosophy of science and has little if any direct connections to reductionism as social scientists conceive it.

A second approach, adopted here, could be called *naturalist*. A naturalist approach takes social science and the philosophy of social science to be continuous. Philosophical issues are not first decided by a priori conceptual analysis and then applied to the social sciences. Issues in the philosophy of social science are in the end empirical issues, though of course being clear on concepts is important. Moreover, social scientists worry about concepts all the time and often do what could be called philosophy of social science. Reductionism in the social sciences on this view then is an issue, or a set of issues, within social science, albeit one to which philosophical

tools may be relevant. Interesting reductionist claims will have to be investigated in the context of specific social explanations, evidence, and theories.

Reductionism Viewed From the Philosophy of Science

One possible form of reductionism in science in general has been described with some precision by philosophers in terms of what might be called *intertheoretic reduction*. It is assumed that there are well-confirmed and clearly statable theories available. A reduction takes place when all the well-confirmed generalizations or explanations of one theory can be derived from another. Such derivations will require some way to relate the basic categories or terms of one theory to the other, since different theories will use different categories or vocabulary. Usually, this translation is thought to come about via what are called *bridge laws*—connecting statements that link each basic category to some category or categories in the reducing theory. So reductionism in biology might have occurred when the notion of a gene, after having been postulated but not understood in biochemical terms until the mid 20th century, was identified with stretches of DNA that code for a protein. Such translations may call for refinements in the theory being reduced: To continue the genetics example, we might find out when we look at the DNA behind heredity traits that what we thought was one gene was the interaction of two separate genes. Such translations are also not necessarily equivalencies of meaning—the ordinary term *gene* did not have DNA as part of its original meaning, since genes were identified long before DNA was understood. Classical cases of such reduction are thought to be the reduction of genetics to molecular biology, as mentioned above, and the reduction of the gas laws (e.g., temperature is proportional to pressure and volume) to the branch of physics called *statistical mechanics,* which applies Newton's laws to the behavior of particles in a gas. By using the bridging link that temperature is the mean kinetic energy of the atoms of the gas, it can be shown that the gas laws follow from Newton's laws. Putting the further details of this philosophical account aside, the basic idea behind it does seem to capture one important sense of reduction: One theory reduces another when it can explain everything the other theory does.

Advocates of Theory Reduction in the Social Sciences

Both philosophers and social scientists have seemingly advocated this kind of reductionism in the social sciences under the guise of what is called *methodological individualism*. It is common to find philosophers and social scientists saying that everything in the social world can be explained in terms of individuals, their beliefs, their actions, and their relations to each other. Much of modern economics advocates such a view, where the paradigm explanation of social phenomena is in terms of individuals maximizing their outcomes in a market given their resources and preferences (rational choice theory). Numerous social scientists have thought that some such picture applies across the social sciences, not just in economics. For example, rational choice social science tries to show how political institutions, norms, and other social phenomena result from the rational choices of individuals, given their situation and circumstances.

Methodological individualists often define or defend their position with the seeming truism that society does not act or exist independently of real human beings. If an institution or organization exists and behaves in various ways (e.g., corporations exist to pursue profits), it is because individual human beings act collectively to do various things. Social influences, such as "the spirit of the times," that act independently of individuals are mystical and unscientific.

Methodological individualists in the social sciences have defined their position in opposition to what they call *holism*. Holism is often put forth as the view that macrosociological entities such as social classes and governments are real, have "emergent properties"—ones that cannot be predicted from knowledge of individual behavior alone—and even have causal powers of their own in that they can influence individuals. Founders of modern sociology such as Émile Durkheim, who talked about "social facts" that he believed were the domain of the social sciences, advocated holist views. Many contemporary social scientists do so as well, often putting the issue as one about the irreducibility of a social structure that constrains individual behavior. In economics, macroeconomics seems to work in terms of supra-individual aggregate entities such as gross domestic product and aggregate demand.

Reductionists in economics would claim that macroeconomics is not legitimate until it has what

they call *microfoundations,* which is sometimes put explicitly as the reductionist idea that macroeconomic results can be derived from the behavior of individuals maximizing their preferences under budget constraints in markets. Social scientists sympathetic to rational choice models in their disciplines similarly ask for microfoundations and put their demand as one of fully explaining social phenomena in terms of individuals acting on their beliefs and desires.

Important work in agent-based modeling in the social sciences also seems to require such individualist understanding of social phenomena. Agent-based modeling uses simulations, mostly computer based, to show how agents following relatively simple rules—their preferences—produce recognizable social phenomena when they interact. As social scientists sometimes put it, the idea is that you have not explained a social phenomenon until you have grown it from individual behavior. A pathbreaking early instance of this stance is found in the work of Thomas Shelling, who showed how residential racial segregation could arise from individuals who had only a very weak preference for living in neighborhoods where they were not a racial minority.

It is important to note that all these versions of individualist reductionism in the social sciences allow and generally make it mandatory that the interactions and relations among individuals are an important part of the explanation of social phenomena. There are some social scientists, however, who advocate an even stronger or demanding form of reductionism that gives relations and interactions among individuals little or no place. This doctrine is sometimes known as *atomism.* For example, social scientists sympathetic to sociobiology—the explanation of social phenomena in terms of traits in humans evolved by natural selection—sometimes advocate atomist ideas. Wars, for example, might be explained as the result of evolved aggressive tendencies of humans—groups of humans commit aggression because they are composed of aggressive individuals.

Potential Obstacles to Reductionism as Theory Reduction

The philosophy of science literature on this version of reductionism has identified several potential obstacles to viewing reductionism as theory reduction. First, social entities such as social classes, organization, and corporations and social processes such as revolutions are potentially what philosophers call "multiply realizable." They are multiply realizable in that a corporation, for example, can be organized in many different ways in terms of the individuals composing them, and while corporations may try to maximize profit, many different kinds of individual behavior may make that happen.

Multiply realized social entities and processes create a problem for theory reduction. The problem is that there does not seem to be one set of individual behaviors, attitudes, and relations that can be used to translate social categories into the categories of explanations solely involving individuals. Explanations such as "corporations seeking to maximize profit in a competitive market results in wages equaling the marginal product of labor" seem to work fine in invoking social entities and processes, but it is not at all clear what set of individual behaviors we can use to define those entities and processes in individualist terms.

The second problem for reductionism in the social sciences as theory reduction is that when we do explain social phenomena in terms of individual behavior we do so by taken as given and unexplained important social processes and entities. To the extent that we take social entities and processes as given or essential to an explanation, the social has not been reduced to individual behavior. For example, an important type of individualist explanation in the social sciences comes from game theory. Game theory models social phenomena as resulting from individuals pursuing their preferences with given resources by adopting strategies concerning how to react to specific behaviors of others. Game theory takes those possible strategies and shows what set of strategies would be an equilibrium—the situation where each individuals response was his or her best given the responses of everyone else—for individuals "playing" against each other.

Such explanations are about individuals, but they are about individuals in a social context. The norms and socialization influencing their preferences are assumed already to exist, as is the division of resources among individuals; the types of individuals—such as owner and worker—are often assumed to be in place. So these are explanations in terms of individuals but in a preexisting institutional context.

What are the consequences of these potential obstacles? They show that a common argument for reductionism in the social sciences that we saw earlier—that social entities are composed of and do not act independently of individuals—may not be

compelling. While it may be true that social entities do not exist and act independently of individuals, this does not tell us anything about our powers to explain. It does not tell us that the categories we use to explain social phenomena have translations into the categories we use to describe individuals—social categories may be multiply realizable. It does not tell us what our explanations of individual behavior are like—they may invoke unexplained social context.

Another consequence of these obstacles is that reductionism in the social sciences is an empirical issue and we need not expect any uniform answer to the question of whether the social is reducible to individual behavior. Maybe in some cases the multiple-realization problem is not real and explanations in terms of individuals are possible without invoking unexplained social context. Maybe in other cases the problems are real and reductionism is an implausible view.

Yet a third consequence is that successful reductionism may be a matter of degree. Some accounts may be more individualist than others in that they presuppose less unexplained social context. For example, we mentioned earlier the simulation-type explanation of residential racial segregation, which shows that it could result simply from the preferences of individuals to live in proximity with others in their racial group. However, it is an empirical issue whether this is a full explanation in terms of individuals. First, preferences to live with one's racial group may in fact result from participation in institutions and social arrangements promoting racial discrimination in general, and so social factors are still needed to explain residential racial segregation. Second, it is an empirical question even whether preferences for one's own race are sufficient to explain residential segregation. There is evidence that banks may discriminate against qualified African American loan applicants, preventing them from buying in prosperous White neighborhoods. If true, this explanation would explain not just in terms of individuals but also in terms of the social organization of society and its institutions.

Reductionism in the Social Sciences and Mechanisms

Another trend in the social sciences and philosophy of science that is in the spirit of reductionism is one that emphasizes mechanisms, in particular mechanisms describing the behavior of individuals. A contemporary approach in sociology that calls itself "analytic sociology" advocates such views, though many other social scientists have at times expressed such views. The basic intuition is that explanations of social phenomena are incomplete until we see how they are brought about by the actions of individuals.

Social scientists and philosophers have put forth different versions of the idea that we need individualist mechanisms, not always keeping the difference entirely clear. If the idea that we need individualist mechanisms means we can explain everything in terms of individuals, then this version of reductionism in the social sciences is not really different from the theory reduction version described above. Some advocates of producing individual mechanisms in the social sciences, however, are clear that they do not imagine the social sciences explaining without appealing to the existence of institutions and social organization. Rather, they view mechanisms as showing how individuals acting within the social constraints provided by the preexisting social organization and structure produce social phenomena—that is, why individuals act the way they do in the social situations they find themselves in. This, more moderate reductionism has been around since the beginning of the study of society by broadly scientific methods. For example, Karl Marx thought that society had to be explained in terms of the economic classes that composed it. Yet he also believed that people make their own history, but not just as they please—they face social constraints.

This more limited kind of reductionism in social science can legitimately be seen as being behind much of current social science. We mentioned game theory and rational choice theory earlier. Both are extremely influential in the social sciences. Both can be seen as explaining the actions of individuals, given the institutional constraints they face. It seems undeniable that such approaches have led to progress in the social sciences, though some social scientists think that they overstate their success, in part because they often rest on unrealistic assumptions about individuals—for example, that they are fully rational.

There are several other approaches in the social sciences that advocate reduction to individual behavior in social context. The traditions in sociology known as *ethnomethodology* and *symbolic interactionism* are very much about giving accounts of individual action in social context, arguing that such

accounts have a central place. Unlike game theory and rational choice theory, they advocate accounts of individual agency that de-emphasize the role of rationality and maximizing in explaining individual behavior, in favor of what they see as a richer framework involving the processes whereby individuals create meaningful social worlds.

Conclusion

Reductionist ideas have had an important influence in the social sciences. They have taken two main forms: as advocating either that all social phenomena can be accounted for solely in terms of individuals or that the social constraints imposed by societal organization and institutions must be linked to the individual behavior they shape. These programs have been popular and productive to different degrees, depending on the social science in question and the type of phenomena being explained.

Harold Kincaid

See also Analytical Sociology and Social Mechanisms; Ethnomethodology; Holism, in the Social Sciences; Individualism, Methodological; Microfoundationalism; Naturalism in Social Science; Realism and Anti-Realism in the Social Sciences; Reduction and the Unity of Science; Social Choice Theory; Social Facts; Symbolic Interactionism

Further Readings

Durkheim, É. (1982). *Rules of sociological method*. New York, NY: Free Press.

Elster, J. (1982). The case for methodological individualism. *Theory and Society, 11*, 453–482.

Fodor, J. (1974). Special sciences and the disunity of science as a working hypothesis. *Synthese, 28*, 77–115.

Hedström, P., & Bearman, P. (Eds.). (2009). *The Oxford handbook of analytic sociology.* Oxford, England: Oxford University Press.

Hedström, P., & Swedberg, R. (1998). *Social mechanisms: An analytic approach to social theory.* Cambridge, England: Cambridge University Press.

Hodgson, G. (2007). Meanings of methodological individualism. *Journal of Economic Methodology, 14*(2), 211–226.

Janssen, M. (1993). *Microfoundations: A critical inquiry.* London, England: Routledge.

Kincaid, H. (1996). *Philosophical foundations of the social sciences: Analyzing controversies in social research*. Cambridge, England: Cambridge University Press.

Kincaid, H. (2008). Individualism versus holism. In S. N. Durlauf & L. E. Blume (Eds.), *The new Palgrave dictionary of economics* (2nd ed.). London, England: Palgrave Macmillan. Retrieved from www.dictionaryofeconomics.com/article?id=pde2008_I000277

Marx, K., & Engels, F. (2008). *The communist manifesto*. Chicago, IL: Pathfinder Press. (Original work published 1848)

Schelling, T. (1969). Models of segregation. *American Economic Review, 59*(2), 488–493.

Udéhn, L. (2001). *Methodological individualism: Background, history and meaning*. New York, NY: Routledge.

Zahle, J. (2003). The individualism-holism debate on intertheoretic reduction and the argument from multiple realization. *Philosophy of the Social Sciences, 33*(1), 77–99.

Reflective Equilibrium

Reflective equilibrium is a philosophical account of method, justification, and theory construction in ethics and political philosophy. John Rawls introduced the term *reflective equilibrium* to depict a way of bringing principles, judgments, and background theories into a state of equilibrium or harmony. The thesis is that justification in ethics and political philosophy occurs through a reflective testing of moral beliefs, moral principles, and theoretical postulates, with the goal of making them as *coherent* as possible. Reflective equilibrium, as a theory, is therefore often referred to as a form of *coherence theory,* or *coherentism*. Proponents of justification by coherence hold that a theory or set of moral beliefs is justified if it maximizes the coherence of the overall set of beliefs that are accepted upon reflective examination.

The goal of a coherence of norms assumes that there are norms to be shown or made coherent. One must start, in moral and political reflection, with a particular body of beliefs that are acceptable initially without argumentative support. According to Rawls's account, method must begin with what he calls *considered judgments*. These moral and political convictions are those in which we have the highest confidence and that we believe to contain the lowest level of bias or distortion of perspective. Considered judgments deserve this status because they are so deeply entrenched in moral thinking that

any morally decent person would accept and act on them. Examples are norms about the wrongness of racial and sexual discrimination, religious intolerance, and political repression. Considered judgments occur at all levels in moral thinking, from those made in notably particular situations (e.g., a compassionate and caring response when a person is ill) to basic moral principles and rights (e.g., human rights).

The process of achieving a state of equilibrium is a process of reflective testing in which one prunes and adjusts one's considered judgments and beliefs to render the whole set of convictions coherent. The resultant moral and political norms can then be tested in a wide variety of previously unexamined circumstances (e.g., in a broad array of circumstances of apparent conflicts of interest never previously examined) to see if incoherent results emerge. If incoherence arises, conflicting norms must be adjusted to the point of coherence.

Bare coherence never provides a sufficient basis for justification, because the body of substantive judgments and principles that cohere could themselves be merely a system of prejudices and ill-considered judgments. Normative views are often wrong not because they are incoherent but because there is no way, when starting from considered moral judgments in shared morality, that one could, through reflective equilibrium, wind up with anything approximating these beliefs. Accordingly, reflective equilibrium is not a pure coherence theory but a process of starting with considered judgments and then engaging in reflective adjustment of norms and beliefs.

The goal of achieving a state of reflective equilibrium is an ideal in which all beliefs fit together coherently, with no residual conflicts or incoherence. This ideal can only be approximated. A stable equilibrium in the full set of one's moral and political beliefs is an unrealistic goal. Moreover, the trimming and repair of beliefs will occur continuously as new situations of the conflict of norms are encountered. However, this ideal is not a utopian theory toward which no progress can be made. Particular moralities (of individuals and groups) are, from this perspective, works continuously in progress, never finished products.

To take an example from the ethics of the distribution of organs for transplantation, imagine that an institution has used and continues to be attracted to two policies, each of which rests on a considered judgment: (1) distribute organs by expected number of years of survival (to maximize the beneficial outcome of the procedure) and (2) distribute organs by using a waiting list (to give every candidate an equal opportunity). These two distributive principles are inconsistent and need to be brought into equilibrium in the institution's policies. Both can be retained in a system of fair distribution if coherent limits are placed on the norms. For example, organs could be distributed by expected years of survival to persons 65 years of age and older, and organs could be distributed by a waiting list for persons 64 years of age and younger. Proponents of such a policy would need to justify and render as specific as possible their reasons for these two different commitments. Such proposals need to be made internally coherent in the system of distribution and also need to be made coherent with all other principles and rules pertaining to distribution, such as norms regarding discrimination against the elderly and fair-payment schemes for expensive medical procedures.

The ideal of maximal coherence of the overall set of relevant beliefs includes not only moral and political norms but also related empirical beliefs and all initial considered judgments. This is a version of so-called wide reflective equilibrium, in which equilibrium occurs after assessment of the strengths and weaknesses of the full body of all relevant and impartially formulated judgments, principles, and theories. Moral and political views to be included are beliefs about particular cases, about rules and principles, about virtue and character, about consequentialist and nonconsequentialist forms of justification, and the like.

Reflective equilibrium may seem to contain a conservative bias, but the method clearly encourages constant improvement and advances through innovative reformulations and testing that will bring about improved coherence. The method insists that even the most basic considered judgments may need to be revised in the process of reaching for equilibrium. No norm is privileged beyond modification.

Tom Beauchamp

See also Collective Values; Equilibrium in Economics and Game Theory; Idealization in Social-Scientific Theories; Social Contract Theories; Social Norms

Further Readings

Arras, J. D. (2007). The way we reason now: Reflective equilibrium in bioethics. In B. Steinbock (Ed.), *The Oxford handbook of bioethics* (pp. 46–71). Oxford, England: Oxford University Press.

Beauchamp, T. L., & Childress, J. F. (2001). *Principles of biomedical ethics* (5th ed.). Oxford, England: Oxford University Press.

Daniels, N. (1996). *Justice and justification: Reflective equilibrium in theory and practice*. Cambridge, England: Cambridge University Press.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press. (See especially 20ff, pp. 46–50, 579–580; in Rev. ed. (1999), 17ff, pp. 40–45, 508–509)

Strong, C. (2010). Theoretical and practical problems with wide reflective equilibrium in bioethics. *Theoretical Medicine and Bioethics, 31,* 123–140.

Reflexivity

This entry introduces the notion of reflexivity of thought and practice and its critical function, reviews the major roles it has played in the history of the social sciences, and presents recent developments of it.

Reflexivity may be understood as the ability of thought to reflect back upon itself. To that extent, it is a condition of intelligence and a defining characteristic of philosophy. When philosophical reflection is concerned with not taking things for granted, it moves into second-order questions through reflections upon what informs the basis of thinking itself. Intelligence is informed by the possibility that one is in error or has falsely identified or misrecognized an object, concept, or experience.

In both cases, reflexivity has a *critical* function via an examination of the apparently self-evident through invoking a simple distinction between the subject (knower) and object (known) as if one were a reflection of the other.

Reflexivity: A Very Brief History

Different philosophical traditions inform how we view the role of reflexivity in thought and practice. Reflexive acts have been understood in terms of their possibility of inducing a heightened state of self-consciousness in the service of self-transformation, while those writing in the pragmatist tradition have examined reflexivity within language acts as the means through which mind, self, and society are linked.

From a social-scientific viewpoint, a reflexive focus helps understand the relationship between social-scientific knowledge and social practice by acknowledging a mutual interdependence between the observer and the observed.

For Max Weber, the practices of social science must replicate the same qualities that the philosopher Immanuel Kant found within the human mind. The social sciences cannot just collect social facts but are reflexive practices in terms of being ideas about ideas, as reflected in Weber's notion of *ideal type,* which serves as an analytic instrument for the ordering of empirical reality.

Alfred Schütz was critical of Weber for failing to see the episodic nature of human conduct in which we find that meaning is the event, or an act is a meaningful process. From this point of view, *Verstehen* ("understanding") is not a method for doing social research but what social scientists should study because the *life-world* exhibits the basis for our primary experiences of the world itself.

The life-world enables orientation through taking its self-evidence, or prereflexive constitution, for granted. These basic "meaning structures" are analytically rearranged by the social sciences, with the consequence that they do not accurately reflect the basis of social relations. Schütz thus argued that social-scientific constructs must satisfy the *postulate of adequacy*—that is, be compatible with the constructs deployed in everyday life.

This study of "lay reflexivity" is not a subjective state of affairs unique to each individual but an intersubjective one obtained through publicly available forms of communication, including language. Harold Garfinkel, the founder of ethnomethodology, took the analysis of these "experience structures" to examine the *situated* and practical manner in which the processes of recognition and production take place in everyday life through its inherently reflexive character. Other social scientists have been critical of ethnomethodology for not challenging the influence of dominant social structures on everyday life.

Feminist-inspired social science, although a diverse body of thought, examines the dynamics between subject and object not from a position of disinterest or disengagement but from being *engaged* in the construction of this relationship. Scientific abstraction glosses over women's experiences in everyday life, the result of which is the production of a "third version" of events that is explicable neither in terms of the subjectivity of the analyst nor in that of the subject herself.

In *standpoint feminism,* women's exclusion from dominant ways of ruling is deployed productively because an analytic focus from this distance uncovers processes and structures that, from a male point of view, appear natural but actually require

explanation. The result is argued to be a more robust, reflexive social science.

Reflexive Practice: Role and Limits

As social science seeks to explain the dynamics and contexts of our lives, reflexivity can be the unbearable grit in the oil of understanding. While we may agree that there is no simple distinction between reality and its construction, it also has the potential to translate into sterility and paralysis through a relativist route.

Postmodernist writings tend to activate differences via a focus upon the textual construction of reality. This functions to expose, or guard against, aspirations to universal knowledge. The French sociologist Pierre Bourdieu sought to avoid a relativism/objectivism divide through a focus upon agency and structure in social life and the role of a reflexive social science in explanation.

We may recognize ambivalence, but that does not relieve practitioners of the need to consider whether that is a characteristic of the world itself, and hence reflected in their practices, or the result of a disciplinary uncertainty that comes from an inward-looking, nonengaged set of practices.

If the latter predominates, the balance tilts in favor of the mode of production and places social science in question; it also ignores the importance of the reception of knowledge and its potential to inform subsequent actions. If the former is in ascendancy, it confuses the domains of social-scientific and everyday practice. It then relieves practitioners and audiences of the effort needed to reach understanding about the world itself and our place within it.

To guard against such possibilities, the French sociologist of science Bruno Latour has introduced the idea of *metareflexivity* and *infrareflexivity* in order that we can talk about the world, not simply focus on our words about the world. Tim May and Beth Perry refer to *endogenous reflexivity,* in which the actions and understandings of researchers contribute to the production of accounts and referential reflexivity, where the production of accounts meet contexts of reception among audiences who are positioned differentially. The movement from endogenous to referential reflexivity may be characterized as one from reflexivity *within* actions to reflexivity *upon* actions, enabling connections to be made between individuals and the social conditions of which they are a part.

Tim May

See also Ethnomethodology; Life-World; Self-Knowledge; Social Construction of Reality; Social Practices; Weber's *Verstehende* Approach

Further Readings

Bourdieu, P. (2004). *Science of science and reflexivity* (R. Nice, Trans.). Cambridge, England: Polity Press.

Garfinkel, H. (1967). *Studies in ethnomethodology.* Englewood Cliffs, NJ: Prentice Hall.

Harding, S., & Hintikka, M. B. (Eds.). (1983). *Discovering reality: Feminist perspectives on epistemology, metaphysics, methodology, and philosophy of science.* London, England: D. Reidel.

Latour, B. (1988). The politics of explanation: An alternative. In S. Woolgar (Ed.), *Knowledge and reflexivity: New frontiers in the sociology of knowledge* (pp. 155–176). London, England: Sage.

Lawson, H. (1986). *Reflexivity: The postmodern predicament.* London, England: Hutchinson.

May, T. (with Perry, B.). (2011). *Social research and reflexivity: Content, consequences and context.* London, England: Sage.

Schutz, A., & Luckmann, T. (1974). *The structures of the life world.* London, England: Heinemann.

Weber, M. (1949). *The methodology of the social sciences* (E. Shils & H. Finch, Trans. & Ed.). Glencoe, IL: Free Press.

Reification

This entry delineates three central aspects of the concept of reification: first, the economic structures that enable reification at an objective level, namely, commodity production and circulation; second, the subjective aspects of reification that give rise to the depoliticization of capitalist social relations; and third, the cognitive dimensions of reification.

The concept of reification (from Latin *res* for "thing" or "object"—hence objectification) is central to the lexicon of the Western Marxist critique of the capitalist mode of production. Reification refers to the unengaged, depoliticized, subjective attitude that individuals are structurally compelled to take toward the socioeconomic structures that characterize capitalist societies. This term is a central term in the philosophy of the social sciences insofar as it establishes a link between economic structures and an experiential and practical stance that individuals take in relation to economic structures.

Reification as an Objective Process of Capitalist Economy

The foundation of the concept of reification is to be found in an analysis of the objective economic structures that characterize capitalist societies. Karl Marx argued that the *commodity* form is the central structural principle of capitalist societies. Although commodities appear to be merely useful objects that satisfy human needs, commodities in fact are more importantly a form of social relation that is unique to capitalist societies. Commodities are both objects of use as well as objects that are exchanged on the market. Marx showed that commodities are a form of social relation that is contained in the form of an object. The *dual* nature of the commodity, which is both a thing and a social relation, makes it extremely difficult to observe the relationship between commodity production and circulation and the class exploitation that it necessitates. The *structural obfuscation* of the social relations inherent in capitalist production is generally referred to as *commodity fetishism*.

Reification at the Subjective Level

The structural, objective obfuscation of the socioeconomic relations of capitalist society generates the problem of reification at a *subjective* level. The concept of reification has been used to refer to the effect that the structural obfuscation of the social relations inherent in the commodity form has on subjective experience in capitalist societies. In capitalist societies, individuals are structurally compelled to take a passive and disengaged stance in relation to commodity production and circulation. Individuals are generally unable to perceive the ways in which the production of commodities, which appear to be mere things that satisfy human needs, actually compels a form of social life based upon *class exploitation* and *depoliticization* of economic relationships.

The Hungarian Marxist philosopher Georg Lukács, perhaps the most influential theorist of reification, emphasized the subjective and political dimensions of the concept. On his reading, reification is the unengaged and spectatorial experiential and practical stance that individuals in capitalist society take toward the constitutive structures of capitalist society. In a society in which the economy exists as a separate, self-grounding and autonomous realm of social life, operating in a way that is seemingly independent of human will, individuals become unable to perceive the ways in which human agency constitutes these structures. The structures appear to be ahistorical and permanent rather than historically specific to capitalism and socially constituted through human agency. Insofar as reification is a subjective stance that is related to a socio-economic form, the concept articulates the ways in which capitalist domination exceeds what is typically understood by the "economic" sphere in the narrow sense. Reification is both an "economic" problem as well as a practical and experiential problem. Moreover, the concept of reification identifies the ways in which depoliticization itself is central to the capitalist mode of production. Without depoliticizing social relations, the reproduction of capitalist social relations could not take place.

Alternative Accounts of Reification

Finally, the concept of reification also has been used to *critique* a particular kind of rationality, namely, *instrumental rationality*. Instrumental rationality is purely focused on ends and means to the exclusion of other forms of cognition. Theorists who have interpreted reification as instrumental rationality have tended to diverge from the aforementioned aspects of the concept. This usage of reification emphasizes the purely *cognitive* dimensions of reification rather than the relationship between cognition and socioeconomic structures. Two influential theorists of the Frankfurt school of critical theory, Theodor Adorno and Max Horkheimer, emphasize reification as a form of cognition. Their work eschewed the historicist tendencies of the early Western Marxist approach to reification and argued that reification is primarily a form of instrumental rationality that is not specifically bound to the commodity form as a mode of socioeconomic structuring.

This alternative approach, emphasizing the detachment of reification from its basis in commodity production and circulation, has also tended to characterize much contemporary discussion of the concept of reification. In the contemporary discussion, many theorists have tended to favor a concept of reification that focuses on the intersubjective and cognitive dimensions of reification alone rather than on the relationship between the disengaged and passive subjective stance of individuals in capitalist society and economic structures.

Anita Chari

See also Alienation: From Philosophy to Sociology; Capitalism; Frankfurt School and Critical Social Theory; Ideology; Marxism and Social/Historical Explanation; Marxist Economics

Further Readings

Adorno, T., & Horkheimer, M. (2007). *Dialectic of enlightenment* (E. Jephcott, Trans.). Stanford, CA: Stanford University Press. (Original work published 1969)

Honneth, A. (2008). *Reification: A new look at an old idea* (M. Jay, Ed.). Oxford, England: Oxford University Press.

Lukacs, G. (1967). *History and class consciousness: Studies in Marxist dialectics* (R. Livingstone, Trans.). Cambridge: MIT Press. (Original work published 1923)

Marx, K. (1976). *Capital: A critique of political economy* (B. Fowkes, Trans.). London, England: Penguin Books. (Original work published 1867)

Relativism and the Social Sciences: From the Sapir-Whorf Hypothesis to Peter Winch

Linguistic or conceptual relativity has been associated with certain theses in the philosophy of social sciences, playing a key role, in particular, in earlier philosophical and methodological discussions of social and cultural anthropology. This kind of linguistic or cognitive relativism and, in general, the relation between language and thought or between linguistic terms and conceptual categories, which was seen as indispensable to social inquiry by antipositivist or antiscientist circles in the philosophy of social sciences, has also been associated with specific readings of Ludwig Wittgenstein's language-games and forms of life.

This entry critically reviews two major instances of this alleged linguistic/conceptual relativism and dispels the resultant erroneous assumptions about relativism in the social sciences that were once dominant in the philosophy of social sciences.

Origins

It is widely assumed that the mid 20th century saw the emergence of two powerful relativist theories that posed a significant threat to the prospects of the social sciences. This assumption underpins the further widely held assumption that these challenges have been rebutted. These assumptions are only half true.

One of the two theories in question was advanced by the American linguist Benjamin Whorf (1897–1941), building on the work of his teacher, the anthropologist-linguist Edward Sapir, and this theory is usually referred to as the *Sapir-Whorf hypothesis*. The second figure often identified as having advanced a relativist theory is the British philosopher Peter Winch (1926–1997), in his book *The Idea of a Social Science and Its Relation to Philosophy* (hereafter, *ISS*), a book that is informed by the approach to philosophy pioneered by Wittgenstein. While Whorf avowedly advanced a theory of linguistic relativism, Winch sought to advance no theses, and, moreover, his writings neither propound nor endorse relativism.

In the following sections, we will first look at the Sapir-Whorf hypothesis and assess some criticisms of that theory. We will then progress to Winch's writings, clarifying his claims about the nature of social studies. We will see that far from endorsing or providing a variant of Whorf's thesis, Winch actually provides one with the resources for a critique of Whorf.

The Sapir-Whorf Hypothesis and Linguistic Relativism

Whorf's theory of linguistic relativism is founded on his claim that speakers of different natural languages, or more precisely natural languages belonging to different families, operate according to different grammatical rules and that these different grammars shape the thoughts of the native speakers of that language. What follows from such a claim is that the same physical evidence does not necessarily force upon a perceiver the same picture of the universe, for the picture is formed by the grammar of the language. Therefore, different languages are said to entail different pictures of the universe irrespective of those pictures' basis in the same physical data. Grammar, for Whorf, is the "shaper of ideas" and that which provides "thought materials." Therefore, Whorf hypothesizes that the nature of reality is relative to the grammar of a natural language.

While Whorf's theory has been subjected to a number of criticisms, some miss their target by misunderstanding the precise nature of the theory. For example, some took Whorf to be making a point about different vocabularies: for example, that language *A* contains numerous words for phenomenon

X, for which only one word exists in language *B*. This was not Whorf's claim and is of little interest to him. Others have criticized Whorf's foundational claim that different natural languages, particularly those belonging to different language families, such as Indo-European and non–Indo-European, have different grammars. This is challenged by those who subscribe to Noam Chomsky's linguistic theory, wherein Chomsky advances the theory of universal grammar. Such critics reject Whorf's relativism by contesting the grammatical pluralism on which his relativism is founded. If Chomsky's theory is true, then this does indeed seem to undermine Whorf's hypothesis. However, one does not have to subscribe to Chomsky's theory of generative grammar and its basis in a theory of an innate universal grammar to undermine Whorf's relativism. Indeed, one might even believe that one can defend Whorf's theory against a Chomskyan criticism by, for example, questioning the validity of Chomsky's theories. However, there is another, alternative, way to undermine Whorf's thesis without appealing to or subscribing to Chomsky's theories; here, Whorf's argument is identified as residing *not* in the claim that different grammars exist in different natural languages but rather in a set of unacknowledged philosophical assumptions that are operative in Whorf's hypothesis. This criticism of Whorf brings to the fore the philosophical assumptions underlying Whorf's claim that grammatical differences *entail* different metaphysics. In other words, rather than focusing on his grammatical pluralism, one might focus on Whorf's claim that one can *read off* a metaphysics from the grammar of a language. For one thing of note, all too often overlooked in discussions of Whorf's theory, is that his relativism is not *entailed* by his observations about the grammar of different languages; nor did Whorf claim it to be so. There is, rather, a philosophical argument operative in Whorf's thesis, which is assessable independently of his grammatical pluralism. A criticism addressed to this aspect of Whorf's thesis is one that seeks to show that Whorf's thesis lacks intelligibility. John W. Cook's critique is an exemplar of this approach.

On close examination, Whorf's claim that different metaphysics can be read off from different grammars can be shown to fall short of his aims, and this can be seen by examining his remarks about his own language: English. A little like the anthropologist who writes of the primitive superstitions of the tribe he is studying, while it remains literally unremarkable to him that he then prays before dinner and kisses a photograph of a loved one before going to sleep for the night, Whorf's writings about the metaphysics he claims to read off Hopi grammar are accompanied by a distinctly superficial depiction of the grammar of his own language. His arguments about grammatical categories determining the metaphysics of the speakers of the language are consequent on his being led astray by the surface grammar of his own language, because where Whorf took himself to be identifying a metaphysics that can be read off the grammar of English (and, when he turned to study that, Hopi), he was rather reading into English a metaphysics that owed much to his own preexisting, underlying metaphysical assumptions about English grammar. To paraphrase John W. Cook, Whorf was not reading off a metaphysics from the grammar of English but rather reading into English a metaphysics not there. For example, Whorf employs the example of the concept of "time" in support of his claim that metaphysics is read off grammar, but his assumptions about the metaphysics of time are simply read off his observation that in English "time" is a noun, and he seems to assume that nouns must correspond to something. Whorf is therefore being led astray by the *surface* (superficial) grammar of "time." When he then takes himself to have read off a metaphysics of time as being composed of "moments," "time slices," or "time flows," he is actually being led astray by his unacknowledged assumption that "time" as a noun in English must correspond to something.

To summarize, the real problem faced by someone who is persuaded by the Sapir-Whorf hypothesis is not that it is based on a now unfashionable theory of grammar, in light of the prominence attained by Chomsky's work. Rather, the real problem is that what had appeared to many (including Whorf) as a thesis founded upon, if not logically forced upon one by, the results of empirical studies of the grammar of natural languages is actually a set of claims emerging from the observer, such as Whorf, who is in the grip of an unacknowledged picture of grammar that leads him to first misrepresent his own language and then proceed to misrepresent those he is studying.

Peter Winch and Relativism as One Consequence of Scientism

It is widely assumed that Peter Winch, at least implicitly, endorsed some version of Whorf's theory in his *ISS*. It is also widely assumed that Winch

propounded a theory of cultural relativism, based on his observations about grammar and rules. Both these assumptions are incorrect.

Winch claimed that social studies should be seen *as* philosophy. As he put the matter early in his book, where science is concerned with the explanation of particular real things and processes, the philosopher is concerned with the nature of reality in general—what counts as "real" or what we mean by "real." On this view, social studies are better understood as philosophy because their questions have central to them discussions about the meaning of that which is under study (e.g., "happiness," which has been the subject of much recent putatively empirical study). Specific normative questions such as "Should the state promote happiness through policy?" or questions demanding answers in the form of social facts such as "Is Britain happier than Norway?" or "Does socialism produce greater happiness than neoliberalism?" are parasitic upon the general question about the nature of happiness. The social scientist must concern herself with the question of what we take happiness to be or, put another way, what counts for us as happiness: Is it a psychological state, and if so, of what sort? Is it related to flourishing as a member of a species with a particular set of needs? Is a happy life necessarily a good life? Can happiness that is gained through the suffering of others be genuine happiness? Do we accept the drug addict's claim to be blissfully happy, following a fix, as *genuine* happiness? Is there a difference between genuine and subjective/apparent/false/synthetic happiness? And if there is, what are our grounds for saying so, and what are our criteria for "genuine happiness"? Are there degrees of happiness, and if so, how might these be measured? These are unavoidable questions, and they are questions that cannot be answered by empirical study or causal explanation. The point is that any attempt at empirical enquiry into happiness cannot bypass the philosophical discussion as to the general nature of happiness.

Winch's critique of certain assumptions in the social sciences is rich and multilayered; it has deep ethical dimensions that became increasingly prominent in his later post-*ISS* writings. Actions are meaningful, and their meaning is inextricably linked to the context and occasion of action. Believing that meaning can be understood through identification of an action's cause is to misunderstand the nature of action and meaning. So if one's conception of science demands that explanations are stated in the form of lawlike generalizations with predictive power, then that demands a noncontextual identification of act tokens, whereby the role of context in the meaning of an action (a specific act token) can be foregone. However, the nature of action is such that the context—the social situation—is intrinsic to its identity. Moreover, an attempt to deny or bypass the meaningful nature of action leads us to a misrepresentation of not only the actions of those whom the social scientist claims to be explaining but of ourselves too, for first and foremost the social scientist is an ordinary social actor. The act of understanding others involves understanding ourselves, and this is a point Winch increasingly sought to emphasize and one that might easily have been directed at Whorf. Failure to understand others through a misunderstanding of the meaningful nature of action results in failure to understand oneself in a manner that one might depict as *bad faith*. Furthermore, the observation that actions are meaningful leads to the recognition of them as open textured. Should one concede that lawlike generalizations might be possible for some general categories of action, it would still not be possible to rely on those generalizations having nonplatitudinous predictive worth because of their open texture.

Conclusion

In conclusion, Whorf sought to advance a relativist hypothesis, and we have seen that this emerged from his own underlying philosophical assumptions, which ultimately meant that he failed to achieve his goal. Winch was making specific observations about the nature of the questions and problems that are dealt with in social studies and how being in thrall to particular pictures of what counts as a valid form of explanation can impair our understanding of those questions and problems. For Winch, arguments for theories of cultural or linguistic relativism ultimately lack intelligibility, and they can usually be shown to emerge from a failure of self-understanding: a sort of bad faith or intellectual hubris. For Winch, relativism is usually a consequence of a subterranean scientism, and it is the tendency to the latter from which he sought to liberate his reader.

Phil Hutchinson

See also Causes Versus Reasons in Action Explanation; Language and Society; Language-Games and Forms of Life; Relativisms and Their Ontologies; Rule Following; Social Anthropology

Further Readings

Cook, J. W. (1978a). Whorf's linguistic relativism I. *Philosophical Investigations. 1*(1), 1–30.

Cook, J. W. (1978b). Whorf's linguistic relativism II. *Philosophical Investigations, 1*(2), 1–37.

Hutchinson, P., Read, R., & Sharrock, W. (2008). *There is no such thing as a social science: In defence of Peter Winch*. Farnham, England: Ashgate.

Lucy, J. A. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge, England: Cambridge University Press.

Lucy, J. A. (1997). Linguistic relativity. *Annual Review of Anthropology, 26,* 291–312.

Pütz, M., & Verspoor, M. (Eds.). (2000). *Explorations in linguistic relativity.* Amsterdam, Netherlands: John Benjamin.

Sapir, E. (1964). Conceptual categories in primitive languages. In D. H. Hymes (Ed.), *Language in culture and society: A reader in linguistics and anthropology* (p. 128). New York, NY: Harper & Row. (Original work published 1931)

Whorf, B. L. (1956). *Language, thought and reality: Selected writings* (J. B. Carroll, Ed.). Cambridge: MIT Press.

Winch, P. (2008). *The idea of a social science and its relation to philosophy.* London, England: Routledge. (Original work published 1958; 2nd ed. 1990, with an introduction by R. Gaita)

Relativism in Scientific Theories

Relativism in the context of scientific theories refers to views that argue that since judgments about truth and falsity are (always) relative to/dependent upon the individual person or culture, the concept of objective truth is either incoherent or useless. Some versions of relativism are heuristically useful. But strong variants, such as claims that scientific theories can never be assessed by reference to universal norms, lead to self-defeating skepticism—what Susan Haack calls the *New Cynicism*.

This entry offers a survey of influential relativist positions.

Cultural Relativism

According to *cultural relativism,* a fundamental methodological precept in anthropology, the beliefs and behaviors that individuals find acceptable depend on the culture to which they belong. At first glance, this claim is hardly news to any seasoned traveler—we get over our initial surprise that some people in Scotland really like haggis and think there is a Loch Ness monster. Cultural relativism serves as a useful reminder of the diversity of social arrangements and as a prophylactic against ethnocentrism. If a certain group confers religious significance on a huge meteorite, we should not blithely cart it off to our museum. On the other hand, we need not and should not conclude that the origin story associated with that rock has equal validity with the scientific account of meteors. The distinction is this: Everyone has a right to form their own opinion, and we should respect that right. But not all opinions are epistemically equal.

Strong versions of cultural relativism also have dubious theoretical consequences for research in social science. If each culture can only be understood in its own terms, would it not follow that the careful ethnographer should refrain from assuming that humans in various cultures share *any* perceptual, cognitive, or psychological characteristics? Should the ethnographer eschew *any* scientific explanations of social behavior that invoke universal generalizations? Yet might not ethnocentrism itself be a universal tendency? Anthropologists systematically utilize concepts such as kinship, ritual, taboo, and alterity when comparing cultures and are well advised to do so. Social scientists should always *be aware* of the influence of cultural differences, but they need not conclude that their research results are necessarily only valid in their own culture.

The Cui Bono Principle

Another form of relativism notes that since individuals' beliefs and behavior reflect their values, we should apply the cui bono principle to knowledge claims. Just as the detective asks who would have a motive to commit a crime, so we should always ask who stands to benefit if a new research result were accepted as reliable. Professional organizations address such concerns through *conflict of interest* regulations. For example, judges must recuse themselves from cases in which they have a financial interest in the verdict. A recusal does not imply that the judge would in fact rule in a prejudiced fashion; as the saying goes, it is also important that justice be *seen* to be done.

Scientific institutions are aware of this problem. Medical journals often ask researchers to report their sources of funding, but the financial incentives

are still skewed. To gain approval from the Food and Drug Administration, pharmaceutical companies need scientific assessments of new drugs, so they pay scientists to do research. Obviously, the companies are hoping for positive results, but so are the scientists! Journals are sometimes reluctant to publish negative results, and this makes it more difficult to get funding for future research, especially from the same company. One partial solution is to have more government involvement. Recently, there has been action to set up a public archive and to require the results of all pharmaceutical research, including studies with negative results and even incomplete studies, to be posted.

The cui bono variant of relativism reinforces the old proverb that the wish is father to the thought. But like any heuristic, it provides no guarantee. Not all hypotheses that serve the interests of the powerful turn out to be false. And even members of cults awaiting the Rapture will accept unwanted evidence when it is dramatic or recurring. Scientific norms are structured to minimize the biases of individual scientists. Examples include the emphasis on double-blind experiments and the recognition awarded to novel results. Once again, a relativist maxim that at first sounds like a sophisticated epistemic principle turns out to have limited applicability.

Incommensurability

Philosophical critiques of logical positivism have also generated relativist stances. The doctrine of the theory-ladenness of observation says that the observation reports used as evidence in science are inevitably couched in language that reflects theories about the world. Even the commonsense claim that there were two cardinals eating suet from my bird feeder presupposes that the flashes of red are not caused by pressure on my eyeball or holographic projections, and it further presupposes that birds have digestive systems, that a human intended to provide food for birds, and so on. None of this seems remarkable because we all share basic concepts about birds.

But what if someone or even several people, perhaps from a different culture, say that two ghosts wearing red robes were eating wafers from a ceremonial altar on the porch? Well, we might say, this claim cannot count as a scientific observation report because it presupposes that there are such things as ghosts. There are many ways in which the dialogue might continue, but the relativist conclusion is simple: All observations are laden with theory. So any scientific conclusions that we may draw depend on, and are relative to, the theoretical conceptual scheme within which we operate. But does this mean that there is no objective/impartial perspective from which we can appraise and compare the truth-value of claims?

Thomas Kuhn's account of scientific revolutions argued that even in physics, competing paradigms were *incommensurable,* so that when Einstein's theory replaced Newtonian mechanics, what really happened was like a Gestalt switch. Since Einsteinian concepts of mass, distance, and time were quite different from Newton's, strictly speaking, one could not compare experimental predictions from the two systems to see which was more accurate. There was no neutral, shared observation language. The implications for social science seemed staggering. If even physicists with their precise mathematical theories ended up talking past each other, what hope was there for rational debate between competing approaches in economics, psychology, or history?

There have been many philosophical critiques of the holistic theory of meaning that underlies the claims about incommensurable paradigms. To give a very simple example, just because there is no simple English equivalent for the French concept of bête noire or the German idea of Schadenfreude does not keep us from figuring out whether they might apply to a given situation. We can develop a metalanguage in which to compare conceptual schemes. And even Kuhn objected to relativistic interpretations of his account of the history of science, while continuing to emphasize that the rational comparison of competing paradigms was a long, complicated process. We can illustrate the general point by refining the Gestalt switch analogy: It may well be the case that there is no objectively true answer to the question as to whether the famous Gestalt sketch represents a duck or a rabbit. But in science, as in ordinary life, we can collect more evidence. We can view the creature from other angles; we can probe it; we can take X rays. There is no reason to think that the duck/rabbit ambiguity will persist after further inquiry.

Political Implications of Relativism

In the past few decades, relativism has often been viewed as a politically progressive epistemological stance. Scholars in Gender and Cultural Studies have pointed out errors and omissions in both the

sciences and the humanities arising from the biases and parochial perspectives of the intellectual elites who constructed these supposedly authoritative accounts of how the world works. There is no question that inquiry benefits from debate among a plurality of viewpoints and that highlighting any distorting effects of culture and concepts, financial interests, ideology, and identity is all to the good. But in an era where many politicians nonchalantly ignore the basic findings of evolutionary biology, climatology, and macroeconomics, the extreme versions of relativism that are sometimes identified with postmodernism, postcolonialism, and feminist epistemology are far from politically progressive. If it really were true that scientific assessments of truth and falsity could never be objective and could never be more than warring opinions, then we would be left with nothing but a clash of civilizations. Relativism provides no warrant for such a conclusion.

Noretta Koertge

See also Feminist Epistemology; Kuhn on Scientific Revolutions and Incommensurability; Objectivity; Observation and Theory-Ladenness; Postmodernism; Relativism and the Social Sciences: From the Sapir-Whorf Hypothesis to Peter Winch; Relativisms and Their Ontologies; Scientific Method; Strong Program in the Sociology of Scientific Knowledge; Value Neutrality in Science

Further Readings

Boghossian, P. (2007). *Fear of knowledge: Against relativism and constructivism.* Oxford, England: Oxford University Press.

Haack, S. (2007). *Defending science-within reason: Between scientism and cynicism.* New York, NY: Prometheus Books.

Koertge, N. (Ed.). (1998). *A house built on sand: Exposing postmodernist myths about science.* Oxford, England: Oxford University Press.

Koertge, N. (2000). New age philosophies of science: Constructivism, feminism and postmodernism. *British Journal of the Philosophy of Science, 51,* 667–683.

Kuhn, T. S. (2000). *The road since structure: Philosophical essays, 1970–1993, with an autobiographical interview.* Chicago, IL: University of Chicago Press.

Searle, J. R. (2009, September 24). Why should you believe it? [Review of the book *Fear of knowledge: Against relativism and constructivism* by P. Boghossian]. *New York Review of Books.* Retrieved from http://www.nybooks.com/articles/archives/2009/sep/24/why-should-you-believe-it/?pagination=false

Sokal, A. D., & Bricmont, J. (1998). *Fashionable nonsense: Postmodern intellectuals' abuse of science.* New York, NY: Picador.

Relativisms and Their Ontologies

This entry presents varieties of relativism, explains the distinction between the ontic and the epistemic points of view, and critically reviews certain modes in which relativism contrasts with certain varieties of its contrary position—absolutism.

Relativism claims that values such as truth, goodness, or beauty are relative to particular reference frames and that no *absolute* overarching standards to *adjudicate* between competing reference frames exist. Many varieties of relativism can be distinguished, depending on the variables that are concealed in this definition.

Varieties of Relativism

One might be relativist with respect to truth but not with respect to goodness or beauty. One might be relativist with respect to goodness but not with respect to beauty or truth. And one might be relativist with respect to beauty but not with respect to truth or goodness. Still other variables in the definition include the ideas of reference frames, domains of inquiry, and levels (epistemic or ontic). Furthermore, relativists deny one or more strands of absolutism, namely, realism, foundationalism, or universalism.

Merely observing a diversity of beliefs or practices does not entail relativism. One might note such differences, for example, without concluding that truth, goodness, or beauty is frame relative. Also, the distinction between relativism and absolutism is not exhaustive. Antirelativism does not entail absolutism, and anti-absolutism does not entail relativism.

Relativists sometimes argue that any would-be absolute standards would reflect the biases of the absolutist's home culture and that one culture should not impose its values on another. Instead, they hold that we should understand other cultures in terms of their beliefs and values. In turn, some

absolutists worry that denying absolute standards would result in arbitrariness, anarchism, nihilism, or something of the sort. With no absolute standards, inquiries would have no worthwhile goal. By ruling out absolute standards, we rule out the possibility of progress in knowledge. Without absolute standards, we could not distinguish between true and false, advanced and backward, moral and immoral, beautiful and ugly.

Reference frames come in many varieties. They include, for example, cultures, communities, tribes, traditions, religions, forms of life, paradigms, conceptual schemes, and other cognates of "contexts." A person may be relativist with respect to conceptual schemes, for example, but not with respect to tribes. A person may be relativist with respect to communities but not with respect to individuals. Some relativisms are *global,* applying their relativism to all reference frames. Others are *local,* applying their relativism only to some reference frames but not others.

One might be relativist with respect to all or only some domains of inquiry. Relativism with respect to the cognitive domain asserts that a statement is true or false (reasonable or unreasonable, justifiable or unjustifiable) relative to a particular reference frame. Relativism with respect to the moral domain asserts that an action is morally right or wrong (good or bad, virtuous or wicked, praiseworthy or blameworthy) relative to a particular reference frame. Similarly, relativism with respect to the aesthetic domain affirms that something is beautiful or ugly (sublime or mundane) relative to a particular reference frame. Domains may be thought to be made rather than found, or found rather than made.

Ontology and Epistemology

Relativism may be taken to apply at the ontic level (pertaining to existence), the epistemic level (pertaining to knowledge), or both. A relativist at the ontic level embraces the relativity of existents to reference frames, such as virtues or vices being relative to a religious tradition. A global relativist at the ontic level affirms that all existents are frame dependent. A local relativist at the ontic level affirms that only some existents are frame dependent. A relativist at the epistemic level embraces the relativity of our knowledge to reference frames. A global relativist at the epistemic level affirms that all our knowledge is frame dependent, and a local relativist at the epistemic level affirms that our knowledge is only sometimes frame dependent.

Yet the distinction between ontic and epistemic levels is contentious, thus complicating the permutations that assume its validity. One might argue, for example, that if we have no direct access to the world-as-it-is-in-itself, we cannot compare our descriptions with the world-as-it-is-in-itself. At best, we have access to the world-as-it-is-in-itself only through some description of it. We can know the world-as-it-is-in-itself only as conceptualized in one way or another. Therefore, even when we seek to compare a description with the world-as-it-is-in-itself, we are comparing a description with the world-as-it-is-in-itself as conceptualized—which is to say, as already described. We are comparing our description with another description. The most that can be established in this way is a relationship between two descriptions, not a relationship between a description and the world-as-it-is-in-itself.

Thus, one might conclude that the notion of the world-as-it-is-in-itself is of no use in inquiry. Correspondingly, one might argue that the distinction between ontic and epistemic levels is of no use in inquiry. If the distinction between the ontic and the epistemic is deemed a distinction without a difference, those varieties of relativism that depend upon it collapse.

Against Absolutism and Its Varieties

Relativism is partly defined by its rejection of absolutism. Thus, relativists oppose different strands of absolutism, which include realism, foundationalism, and universalism. A strong relativism will negate all varieties of absolutism. A weaker relativism will negate only some of them. A strong absolutism will affirm all of its strands. A weaker absolutism will affirm only some of them. The weaker versions of both relativism and absolutism incorporate both relativist and absolutist strands. Thus, certain varieties of absolutism and relativism are compatible. So relativism and absolutism need not be mutually exclusive.

Consider a strand of absolutism that a relativist might oppose: realism. Realism holds that existents hold independent of reference frames. These existents may be cognitive, moral, or aesthetic. Ontically speaking, realism affirms that sticks and stones, for example, exist as such irrespective of reference

frames. At its epistemic level, realism affirms that our knowledge is secured on grounds independent of any reference frame. "Sticks and stones exist" is true because sticks and stones exist irrespective of reference frames. By contrast, anti-realism in its ontic sense holds that existents are frame relative. Accordingly, whether we conceive particular existents as sticks (or stones) or as collections of electrons in space is relative to a reference frame. While sticks and stones as such exist in a reference frame of middle-sized objects, they do not exist as such in a subatomic reference frame. Thus, the nature and number of existents depend on which reference frame we invoke.

A second strand of absolutism that a relativist might oppose is foundationalism. At the ontic level, foundationalism holds that existents in designated domains are reducible to ultimate and irreducible constituents. For example, Democritus held that objects are reducible to atomic constituents, which he took to be ultimate and irreducible. In its global variant, ontic foundationalism holds that existents in all domains are reducible to ultimate and irreducible constituents. In its local variant, ontic foundationalism holds that existents in only some domains are ultimate and irreducible. In turn, at the epistemic level, foundationalism holds that knowledge can be captured by first principles that are incapable of further analysis, that there must be a terminus to any stage of justification, without which there would be an infinite regress. It posits an irreducible self-evident grounding. In its global variant, epistemic foundationalism holds that irreducible first principles of knowledge hold for all domains. In its local variant, epistemic foundationalism holds that ultimate and irreducible grounds of knowledge hold for some but not all domains.

Relativists might also oppose a third strand of absolutism, universalism. At the ontic level, universalists hold that cognitive, moral, or aesthetic existents obtain for all peoples. Such existents are either frame independent or frame dependent. They are either realist or anti-realist. In the moral domain, for example, universalists at the ontic level might hold that human rights exist for all peoples. Nonuniversalists at the ontic level hold that such rights exist only for some or no peoples. At its epistemic level, universalism holds that no barriers of translation exist between peoples. In contrast, nonuniversalists could argue that incommensurability between designated reference frames blocks universalism. Those of shame cultures cannot fully understand those of guilt cultures, and vice versa. At the epistemic level, nonuniversalists may affirm that the search for a universal language is futile. They may affirm that no universally accessible cognitive, moral, or aesthetic principles for all peoples exist.

At the epistemic level, a universalist might counter that even to know that two reference frames are incommensurable requires both of them to be intelligible to someone who can make the comparison—hence universalism. But this argument is invalid. The possibility that some persons can compare pairs of reference frames does not entail that all peoples are capable of making such comparisons. Furthermore, the possibility that such reference frames are comparable does not entail that they may be *adjudicated* according to some absolutist standard.

We may further qualify universalism in foundational terms. Foundational universalism holds that all peoples, for example, can share some common characteristics, such as basic rights by virtue of what—ultimately or irreducibly—it is to be human. Nonfoundational universalism holds that all peoples share common characteristics without presuming that they do so by virtue of what—ultimately or irreducibly—it is to be human. Foundational universalism is a claim of necessity, while nonfoundational universalism is a claim of contingency.

One might urge that foundational universalism explains the contingent fact of the universal instantiation of certain characteristics—for example, that all peoples have some sort of morality or that all peoples share a taboo against incest. Without foundational universalism, one might urge, the commonality of shared characteristics might appear miraculous. But it is an open question whether such commonality without foundational universalism is miraculous. Evolution theory appears to explain the universality of shared characteristics without recourse to foundational universalism. So universalism appears not to require foundationalism.

In sum, realism does not entail foundationalism; frame independence does not entail ultimate or irreducible constituents. Realism does not entail universalism; frame independence may apply locally, not globally. Foundationalism does not entail universalism; foundationalism may apply locally and not globally. Universalism does not entail realism; that which is shareable by all peoples need not be frame independent. Finally, universalism does not

entail foundationalism; that which is shareable by all peoples need not be ultimate or irreducible.

Concluding Remarks

Pertinent variables that qualify varieties of relativism are values (truth, goodness, beauty, etc), reference frames (cultures, traditions, conceptual schemes), domains (cognitive, moral, aesthetic), or levels (ontic, epistemic). Also, a relativist might selectively negate particular strands of absolutism (realism, foundationalism, universalism). The strongest relativist position will be global, asserting itself across the board. The strongest absolutist position will also be global, asserting itself across the board. Weaker varieties of each combine variables in moderating ways. Accordingly, not all varieties of relativism and absolutism need oppose one another. Relativism is not a single doctrine.

Michael Krausz

See also Epistemology; Essentialism; Multiculturalism; Nihilism; Realism and Anti-Realism in the Social Sciences; Relativism and the Social Sciences: From the Sapir-Whorf Hypothesis to Peter Winch; Relativism in Scientific Theories

Further Readings

Boghossian, P. (2006). *Fear of knowledge: Against relativism and constructivism.* Oxford, England: Clarendon Press.

Hales, S. D. (2006). *Relativism and the foundations of philosophy.* Cambridge: MIT Press.

Harré, R., & Krausz, M. (1996). *Varieties of relativism.* Oxford, England: Blackwell.

Hollis, M., & Lukes, S. (Eds.). (1982). *Rationality and relativism.* Oxford, England: Blackwell.

Kellenberger, J. (2001). *Moral relativism, moral diversity, and human relationships.* University Park: Pennsylvania State University Press.

Krausz, M. (Ed.). (2010). *Relativism: A contemporary anthology.* New York, NY: Columbia University Press.

Laudan, L. (1990). *Science and relativism: Some key controversies in the philosophy of science.* Chicago, IL: University of Chicago Press.

Margolis, J. (1991). *The truth about relativism.* Oxford, England: Blackwell.

Norris, C. (1997). *Against relativism: Philosophy of science, deconstruction, and critical theory.* Oxford, England: Blackwell.

O'Grady, P. (2002). *Relativism.* Montreal, Quebec, Canada: McGill-Queens University Press.

Rorty, R. (1991). *Philosophical papers: Vol. 1. Objectivity, relativism, and truth.* Cambridge, England: Cambridge University Press.

Siegel, H. (1987). *Relativism refuted: A critique of contemporary epistemological relativism.* Dordrecht, Netherlands: D. Reidel.

Reputation, in Social Science

Reputation, from the verb *puto* in Latin, meaning "counting"' or "considering," plus the prefix *re-*, which indicates repetition, is the consideration of the value of an agent by other agents based on his or her past actions and creating expectations on the future conduct of that agent. Reputation is a special kind of social information: It is social information about the value of people, systems, and processes that release information. Reputation is the informational trace of our past actions: It is the credibility that an agent or an item earns through repeated interactions. If interactions are repeated, reputation may conventionalize in "seals of approval" or disapproval or in social stigmas.

The notion of reputation in social sciences has been mainly treated in the field of economics. In Adam Smith's liberal social theory, reputation is seen as a way of coordinating activities in a decentralized social space of transactions. According to Smith, in a free society, markets coordinate diffused knowledge in an asymmetrical way: People have a partial view of what other people know and how they will act. Also, given that most transactions occur over a span of time, parties have to trust each other that they will satisfy their reciprocal interest. These informational and temporal asymmetries call for efficient means of storing and retrieving information about possible partners in interactions. Reputation is more than pure information: It is evaluated information—that is, a shortcut of the many judgments and interpretations that people have cumulated about an actor. That is why people are interested in keeping a "good" reputation by signaling their trustworthiness to potential business partners.

In the rational choice tradition, reputation is modeled as a repeated game. These games raise the question of how you can signal your reputation

before any interaction occurs—that is, how you can signal your credibility in the absence of information about your past behavior. This question is studied within a rich body of work that goes under the name of *signaling theory.* Signaling theory aims at solving a fundamental communication problem: Given an interaction in which interests diverge between the two parties, how can a party be certain of the qualities of the other party? Honest signalers will try to signal their good qualities (trustworthiness, accountability, strength), but dishonest signalers will try to do the same by mimicking high-quality signals. Signaling theory may be traced back to the work of the American sociologist Thorstein Veblen. In his *Theory of the Leisure Class,* published in 1899, Veblen explains the display of wealth of the leisure class (luxury, expensive clothes, time-consuming unproductive activities such as sports) as a way of signaling its social position. Important developments of signaling theory go from the study of behavioral ecology to the sociology and the economy of cultural tastes and lifestyles. An agent emits signals in order to make a threat or a promise credible. *Costly signals* and robust signals—that is, signals that are difficult to fake for those who do not possess the signaled quality—are those considered more credible.

The economist George A. Akerlof has shown that quality uncertainty is such a risky feature of markets that reputation is needed: "Seals of reputation" in a market are labels, certifications, guides—that is, all the devices that tend to reduce the informational asymmetry. A rational agent, according to Akerlof, has an interest in embodying these devices in order to compensate the cognitive deficit of the informational asymmetry.

Quality uncertainty and informational asymmetries have become crucial epistemological issues in contemporary, *information-dense* societies. The vast amount of information available on the Internet and on the media makes the problem of reliability and credibility of information a central issue in the management of knowledge. Informational items that do not come with some label or seal of approval from the appropriate communities are lost in the data deluge of the Information Age.

From the evaluator's perspective—that is, from the standpoint of the agent who has to filter information—reputation has an informational value. This has become a prominent issue in Web studies. Given that the structure of the Web is that of a reputational network, in which each link from a page to another can be read as a "vote" from one page to another, a number of algorithmic techniques have been developed to compute the reputation of different entities on the Web: *recommender systems, collaborative filtering,* and *reputation systems.*

Collaborative forms of sharing ratings are also relevant in the study of *collective intelligence.* People do not share information; they share evaluated and classified information that creates a "reputational stream" of shared judgments. The epistemological implications of the massive use of shared ratings in networked societies are huge; relying on other people's judgments and authority challenges our epistemic responsibility. The reasons why we trust collectively filtered ratings about an item or an agent are seldom explored. Choosing a doctor, an academic institution, or a wine is a way of endorsing a tradition of values—a way of filtering information that is not always transparent and legitimate. Notorious biases in social networks—such as the *Matthew effect,* investigated by the sociologist of knowledge Robert Merton, according to which the nodes of a network that are more prominent have more probabilities to earn more reputation—create noise in the way reputation is diffused.

Other biases need further epistemological and cognitive inquiry. For example, people tend to form beliefs in order to acknowledge previously established reputations, such as voting for a certain party because a very well-reputed friend votes for that party. Also, reputations are resilient and may last over time even when the facts of the matters they are supposed to signal are no longer there. For example, the prestige of institutions and corporations may last a long time after their decay.

Reputation is a social commodity that needs to be handled in a scientific way in order to avoid informational cascades, conformism, and the perpetuation of received views.

Gloria Origgi

See also Information Society; Knowledge Society; Social Capital; Social Networks; Trust, Epistemic; Trust, Social

Further Readings

Akerlof, G. (1970). The market for lemons. *Quarterly Journal of Economics, 84*(3), 488–500.

Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. London, England: Routledge & Kegan Paul.

Gambetta, D. (2009). Signalling. In P. Hedström & P. Bearman (Eds.), *Oxford handbook of analytical sociology* (chap. 8). Oxford, England: Oxford University Press.

Krebs, J. R., & Davies, N. B. (1998). *An introduction to behavioural ecology*. Oxford, England: Blackwell.

Origgi, G. (2012). Designing wisdom through the Web: Reputation and the passion of ranking. In H. Landermore & J. Elster (Eds.), *Collective wisdom* (pp. 38–55). Cambridge, England: Cambridge University Press.

Veblen, T. (2007). *The theory of the leisure class*. New York, NY: Oxford University Press. (Original work published 1899)

Retrodiction and the Epistemology of Future Studies

This entry introduces the field of Future Studies, shows its importance for the social sciences, and reviews the epistemological issues underpinning this new field and its key elements. These include the notions of retrodiction, predicting–explaining (a) symmetry, social reflexivity, and the emerging significance of the social context as essential to knowledge production and therefore the significance of an essentially social-epistemological viewpoint.

Future Studies emerged during the second half of the past century as an interdisciplinary field at the boundary between the social sciences and philosophy. The field deals with a large range of themes related to the cognitive, epistemic, and practical challenges posed to the human mind and human action by the future. As such, it represents one of the most interesting and dynamic areas of interaction between philosophy and social-scientific research. From the very beginning, *epistemological* problems were central to the field's agenda: If the future is not the realm of the true or false but the realm of the possible, what are the implications for the ways we understand, define, and assess our predictive efforts? How do we project knowledge from the past and the present to the future? What is the nature of that knowledge? How do we incorporate the problem of the future in our social theories, and what methods should we use in order to make our conjectural knowledge about the future useful for today's decision-making processes?

As one may expect, the very distinctiveness of the field is centered on the assumption of the asymmetry between our knowledge of the past and our knowledge of the future. The epistemology of Future Studies is built on two major themes, identified by two *asymmetries:* (1) between *prediction and retrodiction* and (2) between *explanation and prediction*. The two themes are related but distinct. The benchmark is considered to be the structural dissimilarity between prediction and retrodiction. Their asymmetry has been described and analyzed using, among other things, the notion that there are many more "signs" of past events than of future events, the idea that there is noninferential knowledge of the past but not of the future, or the argument that a predictive inference must be an inference from a sufficient condition, whereas a retrodictive inference is an inference to a necessary condition.

In a typical move for Future Studies, an actional element is introduced along standard lines of epistemological reasoning. If in articulating our criteria of making true (or falsifying) a statement, the accent shifts from discovering, confirming, or disconfirming (whether it is true or false) to the performance of an action that may bring about (or prevent) the event predicted or retrodicted in the statement, then one could identify two types of situations. The first defines the context of retrodiction, while the second, the context of prediction. One can act now in ways that make true some predictions (or falsify them). One can even make a prediction, and by simply publicly uttering it, one may make it true (self-fulfilling prediction/prophesy) or may falsify it (self-defeating prediction/prophesy). In other words, due to "social reflexivity," the truth-value of future tense statements may be influenced and changed. That is not the case with retrodictions (past tense statements). One therefore needs thoroughly different conceptual frameworks to deal with the two types of epistemic situations.

The relevance for Future Studies of this simple observation is hard to exaggerate. Not only does it help set up the boundary conditions and the epistemic limits of the field, but it also (a) draws attention to the essential role social context and processes have to play and (b) by evoking cases of self-fulfilling prophecies, reveals the contours of the key problem of *social reflexivity*. Social reflexivity

situations are extreme and crucial case studies to be used in investigating the specific epistemic nature of futures research. The very idea that context matters and that, by the simple fact of being made public, pieces of knowledge and information may make empirically true or, by contrast, falsify the very predictions that incorporate them is pivotal for the ways in which the field has defined its epistemological identity and agenda.

Irrespective of how one conceptualizes it, the epistemology of Future Studies is using *retrodiction* as a benchmark. Yet its real foundations lie in its opposition to the conventional "covering-law" model asserting the symmetry between prediction and explanation. The main argument is that the extensions of these two concepts only partially overlap. To explain doesn't necessarily mean to be able to predict. To predict doesn't necessarily mean to be able to explain. There is an overlapping area where the two coincide—that is, they function symmetrically, as prescribed by the "covering-law" model. However, the common area contains much more than the covering-law structures. Probabilistic laws, quasi-laws, and genetic causal models of explanation are located in the same overlapping region. Yet they do not share the formal features and the specific mix of predictive and explanatory power of "covering laws." At the same time, the domain outside the overlapping area is also vast. On the explanation side, there are explanatory structures with no or limited predictive power (e.g., historical or evolutionary explanations). On the predictive side, there are correlational, time-series, analogical, and other predictive structures with limited or no explanatory power. Some predictions could be framed on the covering-law or explanatory models. Most could not. A proper theory and epistemology of prediction should deal with the entire range of predictive approaches and models. To try to force the whole lot in one mold is both unrealistic and unfeasible. This conclusion is a keystone of Future Studies.

Despite the inherent diversity and cognitive heterogeneity of *predictions,* there are nevertheless some common and unifying epistemological themes emerging. Two of them are of special significance. The first is that predictive argumentation is not demonstrative but merely evidential. Hence, formal logic–based argumentative structures of the "covering-law" type are indeed inadequate in giving a complete and accurate account of predictive argumentation and practice. If the nature of predictive arguments is evidential, then the epistemology of prediction should be based not on mere formal logic but on a larger theory of argumentation. That means that the efforts to articulate an epistemology of prediction and forecasting should be based at a formal level on a theory of evidentiary argumentation or, even better, on a general theory of argumentation that incorporates both formal and nonformal argumentative models.

The second theme is that the epistemology of prediction should explicitly deal with the intrinsic *social* nature of knowledge and knowledge production and diffusion. This aspect is especially salient in any situation in which, in order to be forecasting relevant, formal and explicit knowledge needs to be bolstered by background knowledge, with its informal, tacit, personal, and social dimensions. The social dimension is critical in yet another way, already mentioned. When making social predictions, social actors and groups are not inert parameters; rather, they have an intrinsic ability to react and adjust to changes in rules, in the environment, or in their own endogenous dynamics. In a word, they display "reflexivity"—feedback, adjustment, and strategic adaptation. Predictions can be easily invalidated if they become public knowledge and if, based on them, social actors revise and change the patterns of their behavior. That means that understanding the social context and processes associated with forecasting practices becomes intrinsically connected to the efforts to understand the epistemic nature and fate of a forecast. The interplay between the prediction and the social process becomes thus epistemologically relevant at more than one single level.

Thus, the whole cluster of these issues makes the anchoring into the social sciences of the epistemology of prediction (and for that matter of Future Studies in general) an unavoidable part of the Future Studies project. A *social-epistemological* approach emerges not as an option but as a necessity.

Paul Dragos Aligica

See also Covering-Law Model; Inferentialism; Prophecy, Self-Fulfilling/Self-Defeating; Reflexivity; Situated Cognition; Social Epistemology; Tacit Knowledge

Further Readings

Aligica, P. D. (2003). Prediction explanation and the epistemology of future studies. *Journal of Policy, Planning and Future Studies, 35*(10), 1027–1040.

Bell, W. (1997). *Foundations of futures studies: Human science for a new era*. Piscataway, NJ: Transaction.
Helmer, O. (1983). *Looking forward: A guide to futures research*. Beverly Hills, CA: Sage.
Rescher, N. (1998). *Predicting the future*. Albany: SUNY Press.

Risk

Risk denotes potential dangers or otherwise unwanted outcomes. Risk can be used to describe dangers generally, but it also has more narrow meanings, first and foremost linked to probabilistic calculus of possible (negative) outcomes. The concept of risk has evolved over time. This entry gives an account of the changing meanings of risk and the prominent political and social role of risk in contemporary society.

The concept of risk is linked in several ways with Enlightenment and modernity. First, in the early Enlightenment period the concept of risk constituted a new way of calculating potential dangers, which among other things served to circumvent a 14th-century papal ban on primitive forms of maritime insurance. Second, risk is a product of novel forms of inductive reasoning that emerged in the Enlightenment and that consisted of inferences about the future based on the past or on knowledge of causation. Third, linking the two former points, probabilistic calculus (itself one of the absolute key discoveries of this epoch) was the concrete means for such calculation of potential, future, dangers.

Thanks to probability calculus, risk, during the 18th and 19th centuries, became a cornerstone not only in insurance but also in modern forms of government, based (besides new notions of nationhood) on bureaucratic procedures, statistical data on the population, and social insurance regimes.

Concepts of risk based on probability calculus have also become cornerstones in a range of other sectors such as technology assessment and finance, where prognostications are sought after. Early in the 20th century, the economist Frank Knight made a famous distinction between risk and uncertainty that hinges on calculability. Risk, according to Knight, is calculable (by means of probability), while uncertainty is not.

Probabilistic risk calculus has later been significantly refined with new actuarial techniques and advances in knowledge of stochastics and complexity in the latter half of the 20th century. However, beginning in the same period, the Knightean distinction has been challenged in various ways. Risk has increasingly become used to describe also Knightean uncertainty. Furthermore, research on human risk perception and attitudes has demonstrated that risks often are subjectively perceived no matter how objectively risk may be calculated.

Not coincidentally, this development has occurred in a period where public insecurity regarding the side effects, or potential catastrophic consequences, of technology has been strongly voiced and became the center of many political conflicts. Much of the analysis in sociology and political science, especially the theory of a *risk society,* has focused on this new politics of risk and uncertainty. Established institutions and authorities (above all science) are challenged and politicized, and existing (national) political institutions struggle both with the transnational nature of many ecological problems and the ambiguity of the by now, in some respects, deauthorized and politicized scientific or expert knowledge. In this context, risk loses its objective basis, and the distinction between risk and uncertainty is conflated. The risk society thesis sees these developments as constituting a new or second modernity.

Research on human risk perception has added further insights into the subjectivity of risk. For example, it has been shown that voluntary risk taking—engaging actively and willingly in risky practices—increases risk acceptance markedly compared with involuntary or passive risk taking. Humans use a variety of heuristic tools for risk assessment, trying to establish a basis for comparison with known risks. However, this is something that often biases risk perception. Specific types of technologies and practices, particularly those with dramatic impacts and therefore high "dread factor," such as nuclear accidents, tend to be assessed as high risk. Cultural values also influence risk perception. For example, individualistic and entrepreneurial people tend to accept higher levels of risk. In later years, several of these research results have been criticized for only explaining marginal levels of variance in risk perceptions. It has been suggested that specific fears and ephemeral emotional affects (in many cases created by organizational climate), rather than deeply rooted cultural values or cognitive heuristics, influence risk

perception. Research also shows that moral values, for instance, those regarding what is "natural," are a highly influential factor. Finally, risk perception is the result of complex social and communicative processes where news media and centrally placed news transmitters and interpreters influence the public perception of risk. Such processes have been described as the social amplification (or de-amplification) of risk.

While subjective perceptions of risk thus have come to feature prominently in research, techniques for quantifying risk have also become more sophisticated and, above all, have become more widespread in industries such as insurance and finance. However, also in relation to these industries (not the least since the 2008 financial crisis), there have been discussions about the merits of standard probabilistic risk calculus. In particular, in finance, there are discussions today of "fat tails" or "black swans"—that is, low-probability, high-impact events that traditional probability calculus tends to discard as highly improbable.

In summation, risk today is both a prominent political problem and a prominent feature in attempts to control and predict. What unites these two forms is arguably a semantic of the future that accentuates a plurality of potentialities (and individual responsibility for the outcome). Under these conditions, both prognostication and uncertainty come to the fore.

Jakob Arnoldi

See also Bayesianism, Recent Uses of; Complexity and the Social Sciences; Induction and Confirmation; Modernity; Probability; Rational Expectations; Retrodiction and the Epistemology of Future Studies; Time, Social Theories of

Further Readings

Arnoldi, J. (2009). *Risk: An introduction.* Cambridge, England: Polity Press.

Beck, U. (1992). *Risk Society: Towards a new modernity.* London, England, Sage. (Original work published in German in 1986)

Hacking, I. (1990). *The taming of chance.* Cambridge, England, Cambridge University Press.

Luhmann, N. (1993) *Risk: A sociological theory* (R. Barrett, Trans.). Berlin, Germany: Walter de Gruyter. (Original work published in German in 1991)

Rule Following

The topic of rule following is closely associated with the work of Ludwig Wittgenstein. It consists of several interrelated issues: What is a rule? In what ways, if any, can rules guide or compel us? What is involved in committing oneself to following a rule? How do we come to follow rules? What role do rules play in speaking a language? Are rules always and essentially social phenomena? What makes it possible to identify which rules other people are following? Does the investigation of a rule-following community call for a different form of social-scientific investigation than the study of entities that do not follow rules? These questions are obviously of great interest to both philosophers and social scientists.

This entry focuses on the ideas of the two highly influential interpreters of Wittgenstein's rule-following considerations: Peter Winch and Saul Kripke. Winch made the issue of rule following central to the philosophy of the social sciences; Kripke's reading of Wittgenstein dominates the contemporary debate.

Rule Following and the Philosophy of the Social Sciences

Winch uses (his interpretation of) Wittgenstein's rule-following considerations to establish a principled divide between the natural sciences, on the one hand, and the social sciences and philosophy, on the other.

A child uses a word correctly if she uses it in the same way in which she has been taught to use it by her teachers. To speak of actions as "correct" or "incorrect" is to invoke a standard, norm, or rule. Thus, when a child learns new words and linguistic structures, she acquires rules for using these linguistic entities. The child's linguistic behavior can be evaluated as correct or incorrect with reference to these rules, and eventually, she is able to do the same kinds of assessment regarding the linguistic behavior of others.

For Winch, the most important feature of Wittgenstein's analysis of rule following is this: In order to determine whether a given individual *I* follows a rule, we need to study not only *I*'s actions but also the reactions of other people in *I*'s community to *I*'s actions. In other words, we cannot claim that *I* follows a certain rule unless the members of

I's community are able to discover which rule is in question. The reason is that creatures that follow rules are creatures that make mistakes. That is, after all, why we evaluate their actions as correct or incorrect. But for talk of mistakes to have a point, it must be possible to establish that individuals have misapplied the rules they intend to follow. Put differently, it must be possible to overrule an individual's own judgment as to what following the rule demands in given circumstances. It must be possible to appeal to something that is external to the individual in question. And this external check is the responses and judgments of other people.

Rule following does not only concern linguistic behavior. For Winch, all meaningful behavior can be evaluated as correct or incorrect and thus is an instance of rule following.

To understand rule following, Winch claims further, is to accept that the methodologies of the natural and the social sciences must be fundamentally different. Both natural and social sciences involve judgments according to which superficially different phenomena are actually the same kind of entity: Ice cubes and steam are both instances of water; moving the king and castling are both instances of making a chess move. Such judgments of identity presuppose communities, since to say that *A* and *B* are the same is to invoke a rule: the rule that the words *ice* and *steam* stand for some of the phenomena that *water* stands for or the rule that the words *king* and *castling* stand for permissible moves in chess. However—and this is the central point—whereas in the case of the natural sciences we have only one set of rules, in the case of the social sciences we have two. In the case of the natural sciences, we only have the rules that the scientists follow. But in the case of the social sciences, we additionally have to reckon with the rules followed by the people ("actors") under investigation. Winch insists that the latter rules are key for the social sciences. In order to understand, say, a religious community, we must learn which phenomena its members count as the same and as different. This is not to deny that social analysts might develop vocabularies that differ from those of their actors. And yet, the latter have a methodological priority. Winch holds that a study of the concepts of other people must be closer to philosophy than to the natural sciences—after all, conceptual analysis is the heartland of philosophical methodology.

Responses

Unsurprisingly, Winch's book had its greatest impact among philosophers of the social sciences. Much of the response was critical. Many readers were unconvinced by Winch's argument for a sharp separation of the natural from the social sciences. They objected that Winch's methodology is conservative and unable to pose critical challenges to the societies under investigation. A critical social science must (be allowed to) transcend the concepts of the actors. Other critics felt that Winch's focus on concepts and meaningful behavior pushed him toward a form of idealism in which phenomena of violence and suppression have no role. In the same vein, Winch has been accused of reducing social relations to concepts.

Rule Following and Private Language

Although Saul Kripke's interpretation of Wittgenstein's rule-following considerations was not intended as a contribution to the philosophy of the social sciences, it has had a significant impact on the field nevertheless. Kripke offers the following reconstruction of Wittgenstein's argument.

Suppose you are now calculating an instance of the scheme $x + y = z$. Assume furthermore that the respective x and y are larger than in any of your previous calculations according to this scheme. To keep things simple, let us stipulate that x is 57 and y is 68 and that you are infallible in your arithmetical skills. Thus, you will calculate as follows: 68 + 57 = 125. This is the correct answer since presumably you mean *addition* by "+"; it is the correct answer since you intend to follow the rule for *addition;* it is the correct answer since you are committed to using "+" in accordance with the addition function. At this point, Kripke introduces a "rule skeptic." This skeptic asks you to justify your belief that you are committed to following the rule for addition. More precisely, the skeptic challenges you to justify your belief that by "+" you mean *addition* rather than some other function, say *quaddition.* (The quaddition function coincides with the addition function as long as x and y are smaller than 57. Otherwise, the quaddition function gives the result of 5. Hence, according to the quaddition function, 57 + 68 = 5.) The skeptic demands that your justification take the form of your identifying a fact about yourself, a fact by virtue of which you meant and mean *addition*

(rather than *quaddition*) by "+," a fact by virtue of which you were and are following the rule for *addition*. The skeptic suspects that there is no such fact and, thus, no fact as to whether you should reply "125" rather than "5" (or any other number).

This skeptical challenge, Kripke claims, lies at the heart of Wittgenstein's rule-following considerations. Furthermore, Kripke argues that Wittgenstein was right to think that the challenge cannot be met: There is no "straight solution" to the skeptical challenge. We cannot cite any fact that would establish that we are committed to using our words in one way rather than another. To establish this point, Kripke discusses a variety of possible responses to the skeptic. Suffice it here to mention just two.

The first response cites your past usage with the plus sign: You have used "+" for the addition function in the past, so surely you use "+" for the addition function now that you are confronted with "57 + 68." Unfortunately, by our initial assumption, you have never before calculated with two numbers as large as 57 and 68. Hence, your past practice is compatible with the hypothesis that in the past by "+" you meant "quaddition."

The second response invokes your dispositions. When faced with the "+" sign, you are disposed to calculate according to the addition function rather than according to the quaddition function. Hence, by "+" you mean addition. Kripke rejects this proposal, among other things, on the grounds that it fails to capture the so-called normativity of meaning. It is part of our ordinary understanding of meaning that to mean something by a word obliges us to use the word in some ways and not in others. If you mean addition by "+," then you ought to respond "125" to "57 + 68 = ___." This "ought" is not captured by talk of what you are disposed to do. And hence, dispositional facts cannot establish that you mean addition rather than quaddition.

But if all proposals regarding meaning-determining facts fail, do we not then have to conclude, with the skeptic, that—despite appearances to the contrary—we never mean anything by our words and never follow any rules? Kripke denies this. On his reading, Wittgenstein has a "skeptical solution" in answer to the skeptical challenge. This solution focuses on what we do when we attribute meaning to others. When we talk of Mary meaning something by a given word, we are not referring to facts about Mary's mind or behavior, facts that determine how Mary uses the given word. Instead, we signal to others in our community in what kinds of interactions Mary is a reliable person. If (in ordinary circumstances) I say of Mary that she means addition by "+," I tell others in my community that under normal circumstances Mary is reliable, for instance, when it comes to calculating restaurant or grocery bills with prices ranging up to 68 units of my currency. Put differently, when I say to you that "Mary means addition by '+,'" my aim is not to say something that is made true by a meaning-determining fact about Mary. My aim is to say something that is appropriate given your interest in, and curiosity about, using Mary as person with whom you might have an interaction that involves numbers up to 68.

If Kripke's Wittgenstein is right, then talk of meaning and rules makes sense only in the linguistic contexts in which one individual recommends another to the members of his community. To speak of rule following and rule meaning outside such social settings would make sense only if we could find a straight answer to the skeptic. And that, Kripke takes himself to have shown, is not possible.

Responses

Kripke's interpretation of Wittgenstein has been criticized from many different directions. Many critics have sought to defend a straight solution to the meaning-skeptical challenge. Often such defenses have involved the claim that meanings and rules are not intrinsically tied to social contexts. Other objectors have tried to show that the facts of meaning are social facts—that is, that the fact that determines you to answer "125" rather than "5" (in answer to "57 + 68 = ___") is your membership in the social institution of calculating according to the addition function. Another important line of attack has focused on Kripke's idea that meaning is inherently normative. Critics have either argued against this idea or tried to show that normativity itself can be understood in natural-scientific categories.

Martin Kusch

See also Causes Versus Reasons in Action Explanation; Explanation Versus Understanding; Language-Games and Forms of Life; *Naturwissenschaften* Versus *Geisteswissenschaften;* Normativism Versus Realism; Relativism and the Social Sciences: From the Sapir-Whorf Hypothesis to Peter Winch; Social Conventions; Social Rules

Further Readings

Bloor, D. (1997). *Wittgenstein: Rules and institutions*. London, England: Routledge.

Kripke, S. (1982). *Wittgenstein on rules and private language*. Cambridge, MA: Harvard University Press.

Kusch, M. (2006). *A sceptical guide to meaning and rules: Defending Kripke's Wittgenstein*. Chesham, England: Acumen.

Pettit, P. (1996). *The common mind: An essay on psychology, society, and politics* (2nd ed.). Oxford, England: Oxford University Press.

Turner, S. P. (2010). *Explaining the normative*. Cambridge, England: Polity Press.

Winch, P. (1958). *The idea of a social science and its relation to philosophy*. London, England: Routledge.

Wittgenstein, L. (2001). *Philosophical investigations* (G. E. M. Anscombe, Trans.). New York, NY: Wiley. (Original work published 1953)

Scheler's Social Person

"Social person" or *Gesamtperson* (collective person) is the German philosopher Max Scheler's (1874–1928) term for the actual and ideal personhood of community. Elaborated from different perspectives and at various points within his writings—notably in *Formalism in Ethics and Non-Formal Ethics of Values* (1913, 1916), *The Nature of Sympathy* (1913), and *Die Wissensformen und die Gesellschaft* (1926)—Scheler contended that a community is also a person, possessed of all qualities of personhood including consciousness and moral responsibility. Social person is similar to conceptions of intersubjectivity associated with the phenomenological tradition and also similar to corporate understandings of community in the tradition of life philosophy.

In Scheler's *The Nature of Sympathy* (a critical analysis of sympathy as the basis for community, friendship, and love), he outlined the theoretical genesis of the person. His account loosely parallels G. W. F. Hegel's conception of the development of consciousness in the *Phenomenology of Spirit* and similar thinking by other German idealists but is couched in Scheler's phenomenological understanding of consciousness as intentionality. Explored in *The Nature of Sympathy* are the bases of our relation to others, such as in community, friendship, and love. Scheler perceived these bases as stages from lowest to highest, moving from emotional identification, to vicarious feeling, to fellow feeling, to sympathy, and finally to love. In this, Scheler maintained that personhood is the completion of a process that begins with a primitive condition of ecstatic experience wherein intentionality is unreflective and immanent (*Mitwelt*). At this point, the self cannot be discerned. Self emerges only subsequently as human beings experience their environment and even other human beings as resistances to intentions. Gradually, this engenders self-awareness and alienation, manifesting at first and immaturely as an egoistic celebration of individuality. Strident, immature individuality then generates ever sharper alienation that begs overcoming the distance between the self and others by love. The mature person, thus, recognizes herself as an aspect of her loving participation in community with others. A community of such loving participation is also a social person.

In *Formalism in Ethics and Non-Formal Ethics of Values,* Scheler's phenomenological study of value ethics, he offered his most detailed discussion of the social person. He argued there that the person is a center or constellation of acts that occur only in the context of a community. In every execution of an act, the person also reflects participation in the community that encompasses her. Community then is an essential feature of a person's acts. From this, Scheler concludes that every person has an individual person and a social person. This means, among other things, that the individual person and the social person are mutually coresponsible and separately self-responsible for whatever ethical implications acts have. As its own person, the social person of the community, like the individual person,

perceives morally weighted values by which its acts are accountable. And, owing to the mutuality of their participation in the other, both the individual person and the social person share in the moral responsibilities of the other. The individual shares in the moral consequences of the acts of the community, and vice versa.

In *Non-Formal Ethics,* Scheler also presented a theoretical account of the social person from the perspective of a typology of community. In form, his account follows the same pattern as seen in his description of the genesis of the person. *Mass* and *herd* are terms for the ecstatic community of contagion in which individuality is not truly present and the community itself lacks self-awareness and responsibility. "Life-community" is what he called the hierarchical, but corporate, community in which individuals can be distinguished by the functions they perform as limbs or members of the whole. "Society" refers to noncorporate associations in which individuals are alienated from others and the whole and cooperation depends upon overlapping self-interests validated by consent and formal rules. *Social person* is Scheler's name for the highest form of community. Social person completes and incorporates the lower community types.

The social person is not a collection of individuals and also not a construct. Rather, from the perspective of the person, it is essentially concomitant with the idea of personhood. The complete person is not possible apart from the social person. Moreover, from the perspective of the community, the social person is also the theoretical ideal of the personhood of community. Scheler called it a solidary realm of love and explained it as an active and ongoing unity of individual persons within an a priori community of persons. Scheler often contrasted his conception of the social person with Ferdinand Tönnies's 1887 analysis of community as *Gemeinschaft* or *Gesellschaft*. For Tönnies, *Gemeinschaft* referred to the thick and vital community of a tribe or village, which he contrasted with the thinner relationships of modern *Gesellschaft,* or "society." A sharper contrast to Scheler's thinking presents itself vis-à-vis Hegel's 1820 dialectic of family, civil society, and state in *Philosophy of Right*. Scheler rejected comparisons of the social person with the state. Conceding that the state is "personlike," he yet insisted that state corresponds best with the "life-community" level of community in his typology, because state lacks the spiritual and loving qualities of the highest form of community.

In Scheler's 1926 study, *Die Wissensformen und die Gesellschaft* (Forms of Knowledge and Society), he spoke at several points about the mind of the social person but did not explore the concept fully. The social person also was an important aspect of Scheler's theological reflections. The loving and spiritual qualities of the union that is the social person, understood both as a community and as a person, were perceived by Scheler to anticipate a totalizing completion in an encompassing loving and spiritual communion with the divine person that is God. In Scheler's late writings, such as *The Human Place in the Cosmos* (1928), he utilized a new analytical perspective based on philosophical anthropology, and a philosophy of history led him to emphasize what he termed the All-man (*Allmensch*) rather than the social person.

Stephen F. Schneck

See also Hermeneutics, Phenomenology, and Meaning; Holism, in the Social Sciences; Intentionality; Intersubjectivity; Self and the Social Sciences; Sociology of Knowledge and Science

Further Readings

Barber, M. (1993). *Guardian of dialogue: Max Scheler's phenomenology, sociology of knowledge, and philosophy of love*. Lewisburg, PA: Bucknell University Press.

Frings, M. (2001). *The mind of Max Scheler* (2nd ed.). Milwaukee, WI: Marquette University Press.

Ranly, E. (1966). *Scheler's phenomenology of community.* The Hague, Netherlands: Martinus Nijhoff.

Scheler, M. (1970). *The nature of sympathy.* New York, NY: Archon Books.

Scheler, M. (1973). *Formalism in ethics and non-formal ethics of values: A new attempt toward the foundation of an ethical personalism.* Evanston, IL: Northwestern University Press.

Scheler, M. (1980). *Problems of a sociology of knowledge.* London, England: Routledge & Kegan Paul.

Schneck, S. (1987). *Person and polis: Max Scheler's personalism and political theory.* Albany: SUNY Press.

Schneck, S. (Ed.). (2002). *Max Scheler's acting persons: New perspectives.* Amsterdam, Netherlands: Rodopi.

Schizophrenia: Psychoanalytic, Phenomenological, and Contemporary Philosophical Approaches

This entry presents the various ways schizophrenia has been approached by psychoanalysis and different schools of philosophy.

The relationship between schizophrenia and philosophy is a topic of great conceptual range and historical depth. The ideas of a number of 20th-century philosophers (Henri Bergson, Edmund Husserl, Martin Heidegger, Michel Foucault, and Ludwig Wittgenstein) have been applied to understanding this psychiatric illness since its recognition around 1900. Beyond this, there is the question of affinity or opposition between two domains. Schizophrenia is the quintessential form of madness; philosophy often stands as the epitome of reason. The relationship between the two has often served as a proxy for the ancient question of the relationship between rationality and its antithesis—the latter conceived variously as unreason, dementia, the passions, or sheer incomprehensibility.

Affinity or Opposition?

The most influential formulation of this relationship in the 20th century is in Michel Foucault's *History of Madness,* a work that associates the advent of the Enlightenment or Age of Reason with a moment in René Descartes's first meditation when he presents madness (persons whose brains are so disordered that they claim to have bodies made of glass) as a condition so devoid of sense as to place the afflicted person wholly outside the charmed circle of mutual comprehension, rendering him unworthy of the philosophical dialogue on which Descartes is embarked.

This vision of madness as a condition of unreason is perhaps the central one in Western thought and has been formulated variously. In Homer and Greek mythology, madness generally appears as the opposite par excellence of rationality, with self-control and clarity of thought usurped by the passions in all their blindness and fury. For Immanuel Kant, madness involved a decline of the higher cognitive faculties of understanding and judgment and failure to work toward true cognition of things. G. W. F. Hegel emphasized social dysfunction involving involution and idiosyncrasy—a regressive inability or refusal to accept the social and linguistic structures that define mind or rationality itself. But though dominant, these visions have not lacked dissenters. In *History of Madness,* Foucault gives pride of place to far more ambivalent visions of what madness might be. He quotes Michel de Montaigne and especially Blaise Pascal's famous statement to the effect that human beings are so necessarily mad that not to be mad would be another form of madness. This suggests that madness is intrinsic to human nature and may even foster insight or imply a paradoxical kind of sanity.

One variant of the dissenting view would assimilate the madman *to* the philosopher, in the sense of finding in him (or her) the truest sources of wisdom and insight, perhaps in association with release from the thrall of authority and convention. In Shakespeare, the apparent nonsense of the fool (or madman) turns out to contain superior insight; King Lear gains wisdom only once he becomes mad. The other facet of dissent likens not the madman to the philosopher but the reverse, thereby showing the errors of too much faith in reason or abstraction. The *ur*-text for our time is Friedrich Nietzsche's *Birth of Tragedy,* a crucial influence on *History of Madness,* which presents Socrates as a literal-minded, even fanatical intellect who ruins the subtle Dionysian/Apollonian mix that accounted for the vitality and wisdom of Greek tragedy.

Madness as unreason, madness as insight, and rationality as containing a kind of madness: These notions form part of our broad cultural heritage. But they also constitute key vectors underlying specific approaches to psychopathology in three traditions in which philosophy and schizophrenia have been brought into dialogue in the past century: psychoanalysis, phenomenology, and Wittgensteinian philosophy.

Psychoanalysis

Sigmund Freud was, for the most part, an advocate of the traditional view of psychosis or insanity, including schizophrenia, as a decline of the rational faculties. Whereas in neurosis the ego suppresses the id, in psychosis it supposedly lets itself be overcome by the id and is detached from reality. Typically, Freud and his followers presented these

developments as involving regression to more primitive levels of personality or stages of development. A different view, more reminiscent of Montaigne or Pascal, was offered by the French psychoanalyst Jacques Lacan.

Lacan views psychosis in general as, in essence, a rejection of what he calls the "symbolic" order or register. The symbolic includes language and all conventional or language-like (diacritical) structures. Hence, the psychotic person's failure/refusal to assimilate this dimension deprives her of the orienting signposts and the symbolic/conceptual capacities enjoyed by nonpsychotic persons. But like all poststructuralists, Lacan emphasizes the *merely* conventional, arbitrary, indeed fictional nature of these symbolic formations, which, though enabling, have no direct correlation with objective reality. Psychosis is therefore viewed as a condition that combines the possibility of true insight (seeing beyond conventional categories, gaining intuition into "the real") with an inability to function in normal society. This dual vision is captured in a famous line from Lacan's seminar of 1973/1974, "Les non-dupes errent": "Those who are *not* duped wander, lost and in error."

Phenomenology

The first phenomenological student of schizophrenia was Eugene Minkowski, a Polish French psychiatrist who studied with the psychiatrist Eugen Bleuler (coiner of the term *schizophrenia*) but was decisively influenced by Henri Bergson, the most influential French philosopher of the early 20th century. Minkowski saw the *trouble genérateur* of such patients as their lack of "vital contact" with the environment and their consequent reliance, instead, on forms of "morbid geometrism" and "morbid rationalism" that lack the vitality, flexibility, and contextual appropriateness of more engaged forms of existence. A congruent formulation was offered several decades later by the German psychiatrist Wolfgang Blankenburg (later elaborated by Giovanni Stanghellini). For Blankenburg, the defining feature of schizophrenia was the loss in such patients of the sense of "natural self-evidence"—the unquestioned obviousness, the unproblematic background quality that normally allows one to take for granted many aspects of the social and practical world. Blankenburg associates natural self-evidence with the "natural attitude" and "everydayness" described by Husserl and Heidegger. He views its loss as having certain affinities with the *epoché* or phenomenological reduction, the setting aside of belief in external reality that is the key method of Husserlian phenomenology. Both Minkowski and Blankenburg describe how loss of "vital contact" or "self-evidence" can eventuate in hyperreflection on experience and the world, a philosophical or pseudophilosophical scrutiny that attempts to compensate for the loss of intuitive grasp.

The most extensive discussion of these aspects of schizophrenia is offered by Louis Sass in a 1992 book. In *Madness and Modernism,* Sass disputes the dementia, regression, and Dionysian visions of schizophrenia to be found in mainstream psychiatry, psychoanalysis, and antipsychiatry. He argues that the mysteries of schizophrenic experience and expression can best be illuminated by comparison with the modernist and postmodernist art, literature, and thought of the 20th century, which is characterized by forms of hyper-self-consciousness ("hyperreflexivity") and alienation. Sass's use of philosophy is extensive. He argues, for example, that the study of delusion in mainstream psychiatry exemplifies the "forgetting of the ontological difference" that Heidegger identifies as the central error of Western philosophical thought. As a result, there has been a failure to recognize the nonliteral *manner* or *way* in which delusional "reality" may be experienced by many patients with schizophrenia—whose experience will often be less literal, less an example of straightforward "poor reality-testing," than is typically assumed. Foucault's analysis of the paradoxes of the modern forms of self-consciousness introduced by Kant (the "empirico-transcendental doublet") is shown to illuminate certain paradoxes central to schizophrenic self-consciousness, which may combine experiences of one's own mind as both the solipsistic center of the All and a mere mechanism subject to causal forces.

These phenomenological perspectives on schizophrenia blend the three organizing vectors mentioned above. They have influenced the psychiatrist Iain McGilchrist's work on brain hemispheres, forms of attention, and cultural–historical parallels, as well as the philosopher Matthew Ratcliffe's analysis of "existential feelings" or "feelings of being"—experiential phenomena that (Ratcliffe argues) demonstrate the impoverishment of the

conceptions of mental life common in analytic philosophy and cognitive science.

Many aspects of the aforementioned, phenomenological perspectives on schizophrenia are synthesized in more operational form in recent work by Louis Sass and Josef Parnas, who conceptualize schizophrenia as a disturbance of "ipseity" or the minimal self—that is, of the most basic sense of existing as a center or subject of experience. Drawing on the work of various philosophers, including Michel Henry and Michael Polanyi as well as Husserl and Maurice Merleau-Ponty, they describe a two-sided self-disturbance. This involves (1) a diminished, basic sense of existing as a subject of experience ("diminished self-affection") together with (2) a complementary tendency toward focal awareness of aspects of consciousness and the body that would normally be experienced in a tacit manner ("hyperreflexivity," basically an "operative," automatic, or nonvolitional process, but one that also occurs in compensatory, even quasi-volitional fashion). Sass and Parnas have argued that this theory is not merely descriptive but has a certain explanatory potential; it has been operationalized by Parnas and colleagues and applied in research on diagnosis and clinical prediction. The closely related, and crucial, issue of disturbed "prereflective" or implicit forms of temporal experience in schizophrenia has been analyzed by Thomas Fuchs.

Wittgenstein

Despite the growing interest in psychopathology (particularly delusions) in contemporary analytic philosophy and cognitive science, little attention has been paid to schizophrenia in particular. The exception is work inspired by Ludwig Wittgenstein.

The most sustained use of a single philosopher can be found in *The Paradoxes of Delusion,* in which Sass applies Wittgenstein to understanding the famous Schreber case. Wittgenstein was a sort of anti-philosopher: a critic of the illusions that philosophy can suffer from when it turns away from standard "forms of life" and "language-games" in favor of abstraction and involutional concerns. Sass uses Wittgenstein in the service of phenomenology, arguing that the delusions of this classic case of paranoid schizophrenia result from a similar disengagement and self-consciousness and demonstrate forms of self-contradiction that are analogous to what Wittgenstein discerns in the philosophical tradition. In this light, Schreber's madness represents a perverse self-apotheosis of the mind: the end point consciousness reaches when, separating from body, emotion, and the social world, it turns in upon itself.

A final philosophical issue involves a questioning of the very possibility of the phenomenological interpretations discussed above, on the grounds that they fail to appreciate the essential bizarreness that is the very essence of schizophrenia. As Angela Woods explains in a book on notions of schizophrenia in psychiatry, psychoanalysis, and cultural theory, schizophrenia is the "sublime object of psychiatry," in the sense of being recognized precisely for its difficult-to-describe, awesome, or incomprehensible status. But is this incomprehensibility absolute and impenetrable? In his classic work *General Psychopathology,* Karl Jaspers (a psychiatrist before he turned to philosophy) actually *defines* schizophrenia as a condition of incomprehensibility that utterly defies empathic comprehension by normal human beings. But many have disagreed; indeed, schizophrenia has been the prime object of phenomenological analysis in psychiatry. The possible recalcitrance of schizophrenia to interpretation and understanding has been explained variously—for example, by arguing that schizophrenic delusions may involve an alteration of what Wittgenstein termed *hinge* or *framework* propositions and thus undermine the very basis of shared understanding or verbal description.

The philosopher Rupert Read has offered a friendly critique of Sass's use of Wittgenstein, arguing that Sass fails to appreciate Wittgenstein's own insistence on the limits of linguistic description, which (in Read's view) runs up against an absolute limit when encountering the doctrine (or pseudodoctrine) of solipsism and the literal *non*-sense of schizophrenic experience and expression. This disagreement turns, in part, on interpretations of the "new Wittgenstein," which are controversial. In reply, Sass criticizes the Neo-Jaspersian polarization of comprehensibility and incomprehensibility in either/or terms, which risks returning to Descartes's rejection of all possibility of comprehending madness. It would be more in the spirit of Montaigne and Pascal to *complicate* our grasp of empathy, namely, to argue (as do both Sass and Mads Henriksen) that although schizophrenia can defy *standard* forms of empathy, it is not closed to more *radical* forms that

recognize the need to postulate a profound alteration of the organizing structures or ontological horizons of human experience, including time, space, and self-experience.

Louis Sass

See also Ego; Existential Phenomenology and the Social Sciences; Foucault's Thought; New Wittgensteinians; Personal Identity and Trauma; Phenomenological Schools of Psychology; Philosophical Psychology, History of; Psychoanalysis, Philosophical Issues in; Therapy, Psychological and Philosophical Issues

Further Readings

Foucault, M. (2006). *History of madness* (J. Murphy & J. Khalfa, Trans.). London, England: Routledge. (Original work published 1961)

Fuchs, T. (2010). Temporality and psychopathology. *Phenomenology and the Cognitive Sciences*. Advance online publication. doi:10.1007/s11097–010–9189–4

Henriksen, M. G. (2011). On incomprehensibility in schizophrenia. *Phenomenology and the Cognitive Sciences*. Advance online publication. doi:10.1007/s11097–010–9194–7

Hoerl, C. (Ed.). (2001). On understanding and explaining schizophrenia [Special issue]. *Philosophy, Psychiatry, & Psychology, 8*(2/3).

McGilchrist, I. (2009). *The master and his emissary: The divided brain and the making of the modern world*. New Haven, CT: Yale University Press.

Parnas, J., & Sass, L. (2011). The structure of self-consciousness in schizophrenia. In S. Gallagher (Ed.), *Oxford handbook of the self* (pp. 521–547). Oxford, England: Oxford University Press.

Ratcliffe, M. (2008). *Feelings of being: Phenomenology, psychiatry, and the sense of reality*. Oxford, England: Oxford University Press.

Read, R. (2001). On approaching schizophrenia through Wittgenstein. *Philosophical Psychology, 14,* 449–475.

Sass, L. (1992). *Madness and modernism: Insanity in the light of modern art, literature, and thought*. New York, NY: Basic Books.

Sass, L. (1994). *The paradoxes of delusion: Wittgenstein, Schreber, and the schizophrenic mind*. Ithaca, NY: Cornell University Press.

Sass, L., & Parnas, J. (2007). Explaining schizophrenia: Relevance of phenomenology. In M. C. Chung, K. W. M. Fulford, & G. Graham (Eds.), *Reconceiving schizophrenia* (pp. 63–96). Oxford, England: Oxford University Press.

Stanghellini, G. (2004). *Disembodied spirits and deanimated bodies*. Oxford, England: Oxford University Press.

Urfer, A. (2001). Phenomenology and psychopathology of schizophrenia: Views of Minkowski. *Philosophy, Psychiatry, & Psychology, 8,* 279–289.

Woods, A. (2011). *The sublime object of psychiatry: Schizophrenia in clinical and cultural theory*. Oxford, England: Oxford University Press.

Science and Ideology

This entry explains the intricate nature of the ideological critique of science. It first charts different notions of ideology and goes on to show how what is commonly taken as the paradigmatic domain of objective truth—scientific knowledge—can be viewed from the standpoint of certain conceptions of ideology as a sociology of knowledge and, furthermore, how certain scientific theories have been criticized as being themselves ideological.

Ideology is one of a family of terms referring to a comprehensive system of thought that is characteristic of an entire society or of a particular socioeconomic class, religious group, political party, social movement, or profession. Other terms in the family are *worldview* (*Weltanschauung*), *belief system, mentality, discourse,* and *paradigm*. These terms are not equivalent to one another, but all describe an abstract foundation of thought and action that is characteristic of a particular social group. Like other terms in the family, the concept of ideology describes a coherent intellectual system that transcends the particular opinions, beliefs, or doctrines that it incorporates and frames. According to many theoretical conceptions, a person can hold (or be held by) an ideology without being fully aware of the way it shapes and inflects perception, thought, and action.

Unlike some of the other terms in the family, *ideology* has negative connotations that are difficult to put out of play when the term is used analytically. Often, when we speak of a doctrine or belief as an ideology, we are suggesting that it is false, highly partisan, or contestable because of its association with particular interests and agendas. Proponents of belief systems that present themselves as universal abhor the term, except when using it to dismiss

doctrines and "special interests" they oppose. Some social scientists use the term in a critical and even denunciatory way, but others try to use it analytically, as a more or less neutral term of art. However, with few exceptions, social scientists do not characterize their own analyses as being ideological.

Genealogy of the Concept of Ideology

Tensions between polemical and analytical treatments of ideology have run through the entire history of the concept. Karl Mannheim, in his landmark *Ideology and Utopia* (1936), traces the word to Napoleon Bonaparte, who used it to denigrate a group of "ideologues" for promoting politically ineffectual ideas during the contentious debates following the revolution. These ideologues were a liberal republican group who promoted a political version of Enlightenment philosophy. One of these so-called ideologues—Destutt de Tracy—actually coined the word *ideology* to refer to a "science of ideas." Already there was a tension between *ad hominem* arguments that dismiss or denounce ideologies by reference to narrow partisan interests and efforts to treat ideology as a variable social or cultural phenomenon. Mannheim took up Destutt de Tracy's initiative, though Mannheim treated ideology not as the name for a science of ideas but as a social phenomenon to be explained in relation to specific historical and existential conditions.

Mannheim employed a historicist analysis in which a particularistic conception of ideology in polemical disputes develops into a general sociology of knowledge. In vulgar (and, not infrequently, academic) usage, the word *ideology* is a polemical weapon—part of a strategy to "unmask" absolutist doctrines that hide the particular interests of those who stand to benefit from them. Especially in Marxist traditions, ideologies are analyzed by reference to class interests, and "false consciousness" is said to result when members of subordinate economic strata adopt, for example, the laissez-faire ideology of the ruling class. Such an ideology is "false" not only because it fails to fit the characteristics of the real world but also because it does not fit the (theoretically imputed) interests of a particular group that holds it. Ideology critique typically involves an effort to unmask objective, apparently disinterested doctrines and cultural practices by revealing the base interests behind them. Mannheim noted, however, that such efforts at unmasking typically imply their own versions of absolutism, thus furnishing grist for further unmasking in an ongoing dispute. When, as in Mannheim's own time during the Weimar Republic in Germany, it should be apparent that all sides in a dispute are playing the unmasking game, it becomes possible to acknowledge that one's own position as well as those of all others can be traced back to particular interests and social backgrounds. Such insight, Mannheim suggested (overoptimistically as it turned out), can then become the basis for an ascent to an analytical position that renounces unmasking in favor of investigating the social sources of all knowledge.

The Sociology of Scientific Knowledge

Detached from the give-and-take of political disputes and the absolutist postures that fuel them, Mannheim's general *nonevaluative conception of ideology* was the cornerstone of his program in the sociology of knowledge. That program involved a set of requirements, which David Bloor later summarized in *Knowledge and Social Imagery* with a scheme of methodological postulates for the sociology of scientific knowledge: *causality* (explaining belief as a function of existential conditions), *reflexivity* (the application of that form of explanation to one's own position), *impartiality* toward the truth or falsity of the beliefs being explained, and *symmetry* (the use of the same general forms of explanation for all beliefs regardless of their truth or falsity). There was a key difference, however, between what Bloor dubbed the "strong program" and Mannheim's sociology of knowledge, which is that the strong program would apply to all scientific and mathematical knowledge without exception.

For Mannheim, a methodological requirement for the sociology of knowledge was to show that a particular body of knowledge does not simply reflect the nature of things or pure logical possibilities. He supposed that the most robust physical laws and mathematical constructions would not meet that requirement, since physical laws, by definition, describe the nature of things and mathematical proofs are the epitome of logical deduction. Bloor, however, made no exceptions. Consistent with Bloor's program, Michael Mulkay reinterpreted Robert K. Merton's norms of science—communalism, universalism, disinterestedness, and

organized skepticism—as rhetorical themes in a professional ideology that promotes extramural governmental support for science, with minimal regulation or public oversight. Thomas Gieryn put forward a similar reinterpretation of the project of demarcating science from pseudoscience or nonscience. Where philosophers searched for demarcation criteria (most famously exemplified by Karl Popper's criterion of falsifiability), Gieryn treated variable historical efforts to demarcate science from nonscience as expressions of a highly successful professional ideology. Demarcation thus became a sociological phenomenon rather than a philosophical criterion. *Boundary work,* as Gieryn called it, is analogous to establishing geographical boundaries in a historical political process. It is rhetorical, pragmatic, interested, and at times ruthlessly indifferent to consistency.

In addition to treating broad characterizations of science as ideologies, social and cultural studies of science followed Thomas Kuhn in *The Structure of Scientific Revolutions* by treating stable complexes of theory and method in the natural sciences as historical *paradigms*. Kuhn likened paradigms to worldviews: holistic conceptions of nature and cosmos that motivate the search for evidence, guide interpretations of data, and infiltrate the analytical language of a science. He argued that revolutionary changes in the sciences do not result from an accumulation of evidence supporting a new paradigm at the expense of the old; instead, supporting evidence often follows a shift in worldview that marks a community's adoption of a new paradigm. Kuhn tried to hold the line against irrationalism and relativism and insisted that he was not suggesting that science was mere ideology. This did not deter others from enlisting Kuhnian ideas in efforts to unmask the objectivity of science to reveal its ideological underpinnings.

Science *as* Ideology

Ideology critiques of science proliferated in the 1980s and 1990s. Some critiques, such as Stephen Jay Gould's *The Mismeasure of Man,* targeted specific historical and contemporary cases, exposing dubious manipulations of data that produced "scientific" support for ideological assumptions about racial, class, and gender differences. Others, such as Richard Lewontin, associated entire developments such as sociobiology and genetic reductionism with social and political ideologies. For the most part, these criticisms adhered to distinctions between valid and biased methodologies and focused on dubious methods and statistical interpretations. Others, particularly proponents of feminist epistemology such as Evelyn Fox Keller, Sandra Harding, and Donna Haraway, went further to launch more general critiques of the social and cultural assumptions that pervade Western science. The critiques went well beyond historical and sociological research that convincingly documented that women were often excluded, explicitly or by more subtle means, from full participation in science and engineering and that those who did participate often received insufficient credit for their achievements. Feminist epistemologists argued that the very contents of science were pervaded by tacit assumptions and metaphors (e.g., DNA as a "master molecule" and nature "herself" as the passive subject of scientific manipulation and control) that privileged a White, Western European, male, executive mentality. Philosophical critiques of the objective/subjective dichotomy—and of "hard" and "soft" sciences—were mapped onto the gender binary and its discriminatory uses. Feminist critiques focused heavily on the life sciences, particularly in areas associated with reproduction and sexual difference, but they were by no means limited to them.

Though often denounced for being part of a broader anti-science movement, feminist and related "standpoint theories" have often claimed to favor a stronger form of objectivity than the one they hoped to dismantle. Instead of being produced within the sealed-off, purified, specialist domain of the laboratory, the hoped-for science would be open to critical voices, inclusive, and relational. Politics would not be treated as external but as a pervasive and explicit part of method. Objectivity would emerge from the conversation among human and nonhuman agents, rather than being treated as the attitudinal and procedural conveyance of "nature itself" into collective human understanding and control.

Critiques of Ideology Critique

The concept of ideology continues to feature in social studies of science, but it also has been subjected to sustained criticism and partial abandonment. The particularistic conception of ideology as an intrusive source of bias that deflects research away from objectivity and value neutrality remains an integral part of arguments both within the sciences and in popular

and academic discussions of the sciences. Far from being antithetical to the ideals of objective science, such arguments often reinforce those ideals by singling out sources of bias and (in the case of commercialized research such as in the pharmaceutical industry) corruption. However, when those ideals are themselves held to be part of a professional ideology—a rhetoric and a politics that secures its authority by denying that it is rhetorical and politically motivated—ideology critique appears to swallow its own foundations. As a regressive explanation that explains explicit expressions and actions by reference to abstract, nonobvious, tacit, and deniable sources of knowledge and action, ideology critique easily spins out of control and becomes indistinguishable from conspiracy theories. Efforts to resurrect social-epistemological grounds in the space vacated by the dismantlement of objectivity often seem convincing only to the converted. Moreover, to speak of "science" as itself an ideology flies in the face of arguments about the disunity of science that are widely accepted in social studies of science. To note this is not to forecast an end to ideology but to suggest that particularistic uses of the concept will remain interesting as a topic, while its generalized uses will remain dubious as an explanatory resource.

Michael E. Lynch

See also Feminist Epistemology; Ideology; Kuhn and Social Science; Pseudoscience; Social Epistemology; Social Studies of Science and Technology; Sociobiology; Sociology of Knowledge and Science; Strong Program in the Sociology of Scientific Knowledge; Value Neutrality in Science

Further Readings

Bloor, D. (1976). *Knowledge and social imagery.* London, England: Routledge & Kegan Paul.

Gieryn, T. F. (1999). *Cultural boundaries of science: Credibility on the line.* Chicago, IL: University of Chicago Press.

Haraway, D. (1991). *Simians, cyborgs and women.* London, England: Routledge.

Harding, S. (2001). Feminist standpoint epistemology. In M. Lederman & I. Bartsch (Eds.), *The gender and science reader* (pp. 145–168). London, England: Routledge.

Latour, B. (2004). Why has critique run out of steam? From matters of fact to matters of concern. *Critical Inquiry, 30,* 225–248.

Lewontin, R. C. (1993). *Biology as ideology: The doctrine of DNA.* New York, NY: HarperCollins.

Lynch, M. (1993). *Scientific practice and ordinary action.* New York, NY: Cambridge University Press.

Merton, R. K. (1973). *The sociology of science.* Chicago, IL: University of Chicago Press.

Mulkay, M. (1979). *Science and the sociology of knowledge.* London, England: George Allen & Unwin.

Turner, S. (1994). *The social theory of practices: Tradition, tacit knowledge and presuppositions.* Cambridge, England: Polity.

Scientific Method

Are there methods of science? And if so, methods for doing what? Most would agree that there are particular methods for observing in each of the sciences, from those employed by, say, an ethologist in observing chimpanzees in Gombe Park in Kenya to those employed by astronomers in observing the positions of celestial objects. There are methods for carrying out experiments, from properly conducting surveys to following the protocols of randomized controlled trials. There are also methods for applying mathematics to some science. And there are methods for reasoning (deductive, inductive, etc.), of the sort we employ in everyday life as well as in science, though science has methods of reasoning that are specific to it.

The history of science has commonly been accompanied by a philosophy of science, an important part of which is a theory of method—the starting point being Aristotle's works. In modern times, Francis Bacon provided in his *Novum Organum* (1620; its full title is *Novum Organum Scientiarum*) some principles of eliminative induction, René Descartes produced his *Rules for the Direction of the Mind* (written in Latin, probably between 1619 and 1628, and published posthumously), and Isaac Newton included in his *Principia* (1687) four "Rules of Reasoning in Philosophy."

Scientific method is commonly taken to pertain to hypotheses (or theories or, more generally, the claims of some body of belief), and this will be the main focus of this entry.

Are there methods for inventing hypotheses (the context of discovery)? Are there methods for

testing them (the context of justification)? If so, are these definitional of science? Some deny that there is such a thing as an autonomous methodology for science. The deniers include sociologists of scientific knowledge (see, e.g., the so-called Strong Program propounded by David Bloor, one of the contributors to this encyclopedia); those such as Michael Polanyi, who thought that scientific methods are tacit insights that do not have an explicit formulation; and, finally, Paul Feyerabend, who famously claimed in his book *Against Method* that "anything goes." This is a position that the later Feyerabend qualified considerably when he said that each move in the game of science may be accounted for by some rule of method, but there may be no overall theory of method of the sort espoused by "rationalists" such as Karl Popper—that is, a view of method that fits all the significant moves that have been made in all the sciences.

Methods of Discovery

Are there methods for discovering hypotheses given some data? Statistics provides many methods for finding a mathematical function to fit, say, a number of data points. Thus, the method of simple linear regression finds a straight line that will be a "best fit" to the data points, where "best fit" can be determined in a number of ways, such as the least sum of the squares of the deviations of the data points from the line. In their 1987 book, *Scientific Discovery,* Pat Langley and colleagues developed several computer programs that generate from data a number of well-known hypotheses in science. For example, their program BACON generates the following laws: From Johann Kepler's own data, it generates his third law of planetary motion (the cube of a planet's average distance from the sun is proportional to the square of the period); from Robert Boyle's data, his pressure/volume law; from Galileo's own data, his law of free fall; and so on. Yet other programs are theory driven and can take into account theoretical assumptions about the data under consideration. It was once a commonplace that there could be no "logic" of scientific discovery, but this quite general claim can no longer be sustained. However, it should be noted that the programs mentioned do not generate laws only out of data; they need to be supplemented with some heuristic principles to guide their search. But the heuristic principles of the programs are not like principles of method as traditionally understood.

The remainder of this entry concerns some of the methods for testing hypotheses, whether of the natural or human sciences. Some suppose that the methodologies of the two areas of science are quite different, especially the explanation of human action as opposed to natural occurrences. However, there are good grounds to suppose that there is a unity of method across all the sciences.

Randomized Controlled (or Clinical) Trials

The randomized controlled (or clinical) trial (RCT) is recognized as the gold standard of scientific method; it has a wide range of uses from medicine to agriculture, education, and psychology. In the simplest kind of RCT applied to humans, a sufficiently large group of people are randomly selected from the wider population and then randomly divided into two groups: (1) the experimental group E, in which some intervention I is to be performed, and (2) a control group C, in which no such intervention is to be performed (or perhaps some "placebo" is given). The intervention I can be some medical treatment or some mode of instruction (say in reading in the case of education), or whatever. Suppose O is some outcome in which an experimenter has an interest. Then one needs to discover the frequency of O in the case of E with I (suppose it is high) and the frequency of O in the case of C without I (suppose O does not occur or occurs with only low frequency). Then one can claim that intervention I is a significant difference maker in that for the E group I brings about the desired outcome O (with a high frequency) while in the C group without I, O does not occur (or with low frequency).

Randomization is important, along with matters such as the size of the sample, if RCTs are to work satisfactorily. An important presupposition is that the members of the two groups, E and C, are sufficiently similar to one another, the only salient difference being the presence of I in E and the absence of I in C. One cannot always be sure that this idealizing presupposition holds. There are some strategies, however, for making the presupposition highly probable. One of these is to ensure that there is no selection bias of those in group E so that, say, more in E have a casually relevant feature that brings about O while those in C do not; random selection

would reduce the chances of such a selection bias arising. Another is to reduce the possibility of the presence of unknown "confounding variables." In the above experiment, we wish to show that I is an "independent variable" that gives rise to the "dependent variable" O; however, one could establish a spurious dependence when there is really no dependence between I and O but some unknown factor X is the cause of them both. To illustrate, a simple experiment can be performed to show that bottle-fed children are less prone to getting diarrhea than those who are breast-fed (contrary to an initial expectation, since dirty bottles may be used). However, this simple experiment is prone to a confounding variable, such as the level of education of the mother. More educated mothers are more likely to bottle-feed and also to adopt more hygienic practices in using the bottle, thereby reducing the incidence of diarrhea. Thus, a confounding variable, the mother's level of education, is related to both bottle feeding and the incidence of diarrhea (which it suppresses). The role of confounding variables can be reduced through not only increasing sample size but also randomization.

Other important aspects of RCT (as applied to humans) concern whether they are single-, double-, or triple-blinded and what ethical protocols are to be followed when experimenting on human subjects. Also, RCTs fit well with a statistical analysis of the data that can be obtained. And refinements need to be made to avoid Type I and Type II errors and to take into account the investigation of events that occur rarely.

The Hypothetico-Deductive Method

The hypothetico-deductive (HD) method is best suited to the examination of hypotheses concerning items that either cannot be observed (e.g., moving tectonic plates) or are unobservable (electrons, the Higgs boson, etc.). The general idea is to examine some hypothesis (or hypotheses) H (the "hypothetico" part) and then draw out some conclusions from H that are open to direct testing (the "deductive" part, but nondeductive inferences can be admitted as well). However, no deduction can be made from H by itself—auxiliaries are needed. To consider this further, let H be the conjunction of Newton's three laws of motion and the law of universal gravitation. These need to be applied to some model M of a real system (e.g., the solar system or a swinging pendulum). Extra information may need to be supplied, such as the initial state of the system at a given time, IS. And finally, some background theory B not under test may be needed to facilitate the deductions (e.g., if the hypothesis under test is about the drag of the atmosphere on a swinging pendulum, then Newton's own theory of motion may need to be supposed, but it is not under test). Granted H, along with auxiliaries M, IS, and B, suppose a test consequence C is deduced. This is to be compared with some evidence E, which may be a single observation, but more likely E is a low-level regularity obtained from the statistical analysis of observational data or even data arising from clinical trials. Either C entails E (in which case all of [H & M & IS & B] pass the test) or E and C are inconsistent (and all of [H & M & IS & B] fail the test).

As can be seen from the above, the HD method does not come equipped with a significant theory of confirmation or disconfirmation. In the first case, if all of [H & M & IS & B] pass a test (since C passes), then this still does not show that H (nested as it is in the context of other hypotheses) is to collect all the confirmatory support due to the one pass; there are yet other tests to be made. It might appear that the problem of induction makes its appearance at this point but in a more indirect form. In the second case, if all of [H & M & IS & B] fail the test (since C fails), then this does *not* show that the failure is due to H itself; one or more of the auxiliaries, M, IS, and B, may be false instead. So direct falsification of H is not possible (unless the auxiliaries have independent high confirmation). This also shows that a version of the so-called Duhem-Quine problem can come to haunt the HD method. Finally, matters can be complicated by the fact that evidence E might be collected on the basis of the supposition of some further background theory K, such as a theory about the instruments being used to generate E. For example, radiocarbon dating is used to test rival theories in archaeology; but such dating presupposes several background hypotheses K about the presence of radioactive carbon that are themselves theoretical in nature.

There is also a problem with the HD method that bothers the logically minded; this is the problem of irrelevant tacking. Suppose that H (with whatever auxiliaries A) leads to conclusion C, which passes a test. Then any other irrelevant proposition X, when

tacked onto H, as in the conjunction H & X (along with the same auxiliaries), will also lead to the same conclusion C. So the irrelevant X piggybacks its way onto whatever confirmation accrues to H. This is regarded as an undesirable consequence that needs to be remedied, but it is not easy to see what the solution is. These issues aside, the HD method is widely used in some of the sciences, for example, volcanology and tectonic plate theory: Hypotheses about unseen activity under the surface of the Earth are applied so they have testable consequences.

Karl Popper's Critical Rationalism

Popper is a hypothetico-deductivist who exploited certain features of the HD method to set up his own account of science. Popper's notion of falsifiability, or better testability, arises straight from the HD model. For a hypothesis H to be scientific at a given time (i.e., open to test at a time) it is required that it have some testable consequences (at that time); if it lacks any, then it is not scientific (though it is still meaningful). This is the case for the ancient Greek speculative hypothesis that all matter is made of atoms. For most of the life of this hypothesis, it was not open to test since it lacked the necessary auxiliaries that, when conjoined to H, yielded testable consequences. Only since the beginning of the 20th century, as, for example, in Einstein's 1905 paper on Brownian motion, were the right auxiliaries available to provide a test consequence, in this case a lower-level law that could be tested against the zigzag motion of the Brownian particle. Again, if a theory has been falsified, then it must be falsifiable. This allows many claims from the past that intuitively we would now not regard as scientific at all, such as the spontaneous generation of life from matter. But this is open to test and has been shown to be false.

For Popper, the distinguishing feature of science is its ability to constantly revise itself. Popper's proposal that the claims of science be open to test is meant to capture this distinguishing feature. In fact Popperian falsifiability, and so susceptibility to revision, is an aim of science, an aim that the rules of the method of science are intended to realize. To this end, Popper proposed in his *Logic of Scientific Discovery* a number of rules of method the foremost among which is the supreme meta-rule: "The other rules of scientific procedure must be designed in such as way that that they do not protect any statement in science against falsification" (sec. 11). Following from this are his three anti–ad hoc rules, the first of which concerns theory revision through the introduction of new auxiliaries and states, "Only those [auxiliary hypotheses] are acceptable whose introduction does not diminish the degree of falsifiability or testability of the system in question but, on the contrary, increases it" (sec. 20). This anti–ad hoc rule is not without its critics, such as Paul Feyerabend and Adolf Grunbaum. Popper invites us to think of the rules of method as hypothetical conditionals ("if-then" sentences) that say that following rule r will realize value v (e.g., falsifiability) all the time (or with high frequency)—that is, if methodological rule r is followed, then scientific value v is realized. Such a conception of methodology as rule–value pairs is more fully developed (but in a different direction from Popper) by Larry Laudan in his *Normative Naturalism*.

Bayesianism and Probabilistic Methods

If there is a leading theory of scientific method today, it is the probabilistic theory of reasoning embodied in Bayesianism. This contains two fundamental ideas. First, in its simplest form, is Bayes's theorem: $p(H, E) = p(E, H) \times p(H)/p(E)$, where H is some hypothesis and E some evidence, and "$p(\ldots,—)$" is the two-place probability function (the theorem can come in many other interesting forms). The other principle is that of conditionalization and tells us how we ought to adjust our probabilities in light of the new evidence, E, we have learned. It says that $p_{final}(H) = p_{initial}(H, E)$. This captures the following simple inference. Suppose, for example, that the probability is high that the creek running past your house will overflow given two days of steady rain but in the absence of steady rain the probability of its overflowing is low. Then you learn that there have been two days of steady rain. So what ought your (absolute) probability that the creek is overflowing be? The intuition that it is high fits the principle of conditionalization.

Bayes's theorem is able to capture the better aspects of the HD method while avoiding the problem of irrelevant conjunction (tacking, as we saw above); and in addition it comes equipped with an account of confirmation that the HD method lacks. Interestingly, all the principles of method espoused by the later Thomas Kuhn (his model of theory

choice as a set of weighted values) have an account within Bayesian methodology. And obviously, Bayesianism fits within the probabilistic inferences of science. In fact, a full probabilistic account of method will begin with the axioms of probability from which Bayes's theorem follows.

One way to understand Bayesianism is to interpret the probability function p(H) to be something like a person's degree of rational belief in H. Here, the requirement of rationality is simply that the person distribute her degrees of belief in accordance with the probability calculus. This leads to the interesting, so-called Dutch Book Theorem, which shows that if one does not distribute one's beliefs in this way then one can lose in bets against nature. A similar independent proof of the rationality of conditionalization is also available. Also of interest are certain "convergence" results. Suppose two scientists propose quite different initial prior probabilities for the same hypothesis; then it can be shown that as evidence comes in, these differences get swamped and the scientists come to agree on what the conditional probability of the hypothesis is given the body of evidence.

Robert Nola

See also Abduction and Inference to the Best Explanation; Bayesianism, Recent Uses of; Causal Explanation, in Philosophy of Science; Covering-Law Model; Deduction; Explanation, Theories of; Falsifiability; Feyerabend, Critique of Rationality in Science; Hypothetico-Deductivism; Induction and Confirmation; Popper's Philosophy of Science; Probability; Serendipity

Further Readings

Friedman, L. M., Furberg, C. D., & DeMets, D. L. (1998). *Fundamentals of clinical trials* (3rd ed.). New York, NY: Springer.

Gower, B. (1997). *Scientific method: An historical and philosophical introduction.* London, England: Routledge.

Howson, C., & Urbach, P. (2006). *Scientific reasoning* (3rd ed.). Chicago, IL: Open Court.

Langley, P., Simon, H., Bradshaw, G., & Zytkov, J. (1987). *Scientific discovery: Computational explorations of the creative process.* Cambridge: MIT Press.

Laudan, L. (1996). *Beyond positivism and relativism: Theory, method and evidence.* Boulder, CO: Westview.

Losee, J. (2001). *An historical introduction to the philosophy of science* (4th ed.). Oxford, England: Oxford University Press.

Nola, R., & Sankey, H. (Eds.). (2000). *After Popper, Kuhn and Feyerabend: Recent issues in theories of scientific method.* Dordrecht, Netherlands: Kluwer Academic.

Nola, R., & Sankey, H. (2007). *Theories of scientific method.* Durham, England: Acumen.

Popper, K. (1959). *The logic of scientific discovery.* London, England: Hutchinson.

Salmon, W. (2006). *Reality and rationality.* Oxford, England: Oxford University Press.

Scottish Enlightenment: Influence on the Social Sciences

Eighteenth-century Scotland witnessed a period of tremendous intellectual achievement and discovery as well as economic growth. The Scottish Enlightenment—centered in Edinburgh, Glasgow, and Aberdeen—ushered forth developments in the fields we now regard as history, sociology, psychology, economics, and philosophy, as well as in geology, chemistry, medicine, and architecture. Such varied accomplishments ensured that the Scots, long situated on the geographic and intellectual margins of Europe, were now at the forefront of thought.

Many Scottish thinkers had a primary and fundamental interest in the development of society. Some of their concerns may have arisen from the peculiar history of their nation. The 1707 union between Scotland and England generated economic growth, even as Scotland remained divided between the more advanced Lowland and the Highland culture. A variety of Scottish thinkers sought not only to explain social development but to explore the emergence of particular phenomena—language, law, morals, and social conventions, along with trade and economic growth. Although the work of these thinkers did not bring forth, in any direct way, the disciplines of social science that we now enjoy, their endeavors generated lines of inquiry that were original and influential as well as systematic and empirical.

This entry summarizes some of the intellectual background to the Scottish Enlightenment, examines specific areas in which Scottish thinkers made salient contributions, and focuses on several thinkers of significance.

The Background to the 18th Century

Several thinkers were particularly influential to the Scottish study of society. In his *Leviathan* (1651), Thomas Hobbes (1588–1679) developed an account of social and political order by describing what he took to be the natural properties and tendencies of the human being. Hobbes drew inspiration from Galileo Galilei (1564–1642), whose method of resolution and composition Hobbes deployed to resolve society into its smallest components—the individual and his or her particular qualities. Having characterized the human being as manifesting a seemingly egoistic drive to quench desire after desire, Hobbes explained, via a series of purportedly deductive arguments, how individuals so described would bring about *either* conflict, suspicion, and war *or,* alternatively, peaceful cooperation via the institution of a governing authority. In utilizing naturalistic description to deduce regularities of behavior and interaction, Hobbes delineated the development from a state of nature to a law-governed society. His appeal to a scientific description had enormous weight, but other thinkers also proved influential to the Scots. In *On the Law of Nature and of Nations* (1672), Samuel Pufendorf (1632–1694) argued that the human being possesses social as well as egoistic proclivities. His outline of the progress of the human being from a primitive, savage state to a developed and complex society would resurface in the works of several Scots.

The notion that society might develop from a more primitive to a more advanced state was also articulated by Bernard Mandeville (1670–1733), who rejected the idea of social contract and articulated a thesis of unintended social outcomes, as well as a naturalistic account of the origin of morals and society (*The Fable of the Bees,* 1714, a two-volume work that included the essay "An Enquiry Into the Origin of Moral Virtue"). By the middle of the 18th century, however, a work appeared of signal importance to the Scots and to the development of social science: *Spirit of the Laws* (1748), by Baron de Montesquieu (1689–1755). In reflections on the interrelated nature of society and politics, Montesquieu observed how nations differ in circumstance (climate, geography, resources, etc.) as well as in manners and beliefs (religion, law, and cultural mores). He then sought to delineate principles that would account for the differential development of societies (each characterized by a "spirit"). In so doing, he examined how some political constitutions are appropriate for distinct societies and how law is determined less by the discerning eye of reason than by forces of circumstance and belief.

Themes

The Scottish Enlightenment included historians, political economists, ministers, and philosophers, many of whom were friends or acquaintances. However, these thinkers did not form a school; neither did their work represent a single unified outlook. Nonetheless, from their varied treatises emerge several coherent themes, many reflecting the influence of the thinkers just noted.

Human Nature. For the Enlightenment Scots, it is hard to underestimate the importance of a seemingly simple idea: The essential basis or starting point of inquiry must be human nature. An understanding of human nature could invoke self-reflection (as it did for Hobbes), but it should also include observation as well as comparison. The relevant observations could be drawn from classical literature or current society, but the reports and testimonies of those who had voyaged to distant lands proved essential as well. Although the Scots recognized, as did Montesquieu, the wide divergences among societies, most assumed that human nature was universally uniform. The appeal to uniformity had implications both practical and explanatory. If variations among and between humans do not reflect essential differences of kind, then neither should one's moral estimation of the humanity of distinct persons. Moreover, any differences demonstrated the necessity of explanation. Since humans shared a basic and knowable nature, the explanation of sometimes stark differences required attention to the varying circumstances of the natural and social environment. In this way, the Scots drew inspiration from Montesquieu, who had referred to how "physical" and "moral" causes might explain the divergences among societies. Against the views of Hobbes and Mandeville, the Scots typically portrayed the human being as a creature of plural motives, some benevolent, some oriented to self. The human being was a creature of reason *and* feeling. Standards of moral judgment, therefore, were not understood either as products of divine revelation or as a priori propositions comprehensible by reason. A person's moral capacity was

identified less with reason than with a moral sense or particular kind of sentiment, one that often reflected a significant social element, as witnessed in David Hume or Adam Smith's distinct appeals to the role of sympathy in generating moral judgments within society.

Naturalism. Their accounts of human nature led the Scots to develop *naturalistic* theories of society and social development. These theories were inspired by Newton's account of the natural world as systematic and law governed. The human being, understood as part of a unified natural system, possesses specific qualities and tendencies original to his nature. An explanation of social patterns and institutions should not partake of theological assumptions (even as the Scots were not as secular or as skeptical as some of their counterparts in Europe) or appeal to final causes, though some (e.g., Adam Ferguson and Adam Smith) did not wholly depart from a broad providentialism. Nor did these thinkers invoke any founding legislators or wise men to explain the progress of society. In fact, many of their accounts predicate that the institutions and conventions of society reflect unintended and cumulative processes of development over time.

Insofar as their theories relied on the assumption not only of a universal human nature but also of a unified system of nature extending from the natural world around us to the realm within, then so did their accounts of human action rely on an empirical and causal understanding of motives and mind. The Scots generally opposed the notion that a priori reason might provide substantive notions of the natural or moral world. The phenomena of the social and psychological world (even the religious, as witnessed by David Hume's *The Natural History of Religion,* 1757) are to be understood in terms of general principles and causal laws of the sort previously utilized for the external world of nature. In society, as in nature, the same types of causes will produce the same kinds of effects, said Hume. For example, the passions of a person work as causes that generate actions. An understanding of these passions would require further causal explanation of a psychological nature. In large groups, the regular sequencing of motives and actions allows, Hume suggested, for the formation of general explanations of human activity. Apart from the appeal to causal explanation, Scottish thinkers also sought to devise social laws that would inform our comprehension of, say, politics (David Hume) or economics (Adam Smith). Although these laws did not carry explicit formulations setting forth conditions of application, the regularities advanced by the Scots implied both explanatory and predictive functions, thereby anticipating the sorts of laws invoked by later social scientists.

Social Change. To understand society is to comprehend its development over time. Having recognized that societies differ across space and time, Scottish thinkers explored the sequential development of society writ large rather than the progress of any particular society. Most assumed that the idea of a founding social contract was without basis in fact or logic, so the Scots sought alternative conceptions. For the Scots, the human being is typically found in groups. Adam Ferguson, Adam Smith, and John Millar were among those who, following Pufendorf and Montesquieu, offered stadial (or stage) theories of the progress of society. The stages of society were typically three or four. For example, Smith delineated a savage society of hunters and fishermen, then a pastoral age of shepherds, followed by an agricultural age, and then a polished commercial age with division of labor, clear rules of law and justice, and a standard of living higher than ever before. Each age was defined in large part by specific modes of production that influenced manners, morals, and law. The transition from one stage to another was not inevitable—and the specifics of the transitions were often vague—but *if* development occurred, then it occurred in this way. An appeal to sequential development is an idea that still resonates in anthropological and sociological studies.

In their accounts of change, the Scots appealed to the actions of individuals, but they also relied on institutions and social rules, including habits and customs. Moreover, social development, either on the large scale of society or in terms of specific laws or institutions, did not rest on the designs and intentions of individuals or great men. Individuals act with regard to local circumstances and with an immediate rather than long-term intention; as individuals act, the aggregative result of their endeavors appears to be a product of design, even though it is not wrought by anyone's intention or through any sort of agreement.

In pointing out the cumulative processes of development, the Scottish thinkers offered accounts that anticipate social evolutionary theories.

Conjectural History. The interest in the progress of society suggests a historical bent to Scottish theorizing. Some who wrote histories—including William Robertson (1721–1793), *History of Scotland* (1759); David Hume, *The History of England* (1778); and Adam Ferguson, *History of the Progress and Termination of the Roman Republic* (1783)—chronicled the development of a specific nation or epoch. Alongside these empirical histories, the Scots also employed a form of history that Dugald Stewart (1753–1828) referred to as "conjectural." A conjectural history offers a hypothetical and naturalistic account of what might have happened over a period of time. In the absence of evidence, but with a knowledge of human nature and the general characteristics of the world, one might hypothesize about the development of language (Adam Smith, *Considerations Concerning the First Formation of Languages,* 1761), the origin of rules of justice (Hume, *A Treatise of Human Nature,* 1739–1740), or the general progress of society (Adam Ferguson, *An Essay on the History of Civil Society,* 1767). Such histories should avoid speculation while nonetheless narrating the causal links between a known later state of affairs and some earlier and relatively unknown age. Histories such as these tended to point out the degree to which human achievement was often the result of the cumulative effort of many persons acting with specific and local intentions.

Economy and Society. A variety of thinkers, including David Hume, Adam Smith, and John Millar, took a great interest in economic questions. Certainly, Bernard Mandeville had recognized both the advantages of the division of labor and the need to set in place a legal framework for growth and prosperity, but the 18th-century thinkers highlighted the importance of specialization for a burgeoning economy and pioneered a model of a self-maintaining system of trade and production. Some, such as Adam Ferguson and Adam Smith, recognized how a division of labor might have unintended consequences disadvantageous to individuals and to society. Economic activity existed alongside other forms of endeavor and was, in fact, embedded within a larger moral and cultural context. Nonetheless, manufacture and trade could be comprehended in a systematic way and shown to have a means of self-coordination not dependent on a wise legislator. David Hume argued that money was distinct from wealth and, in anticipation of contemporary monetarists, noted the significance of fluctuations in the quantity of money ("Of Money," 1752, *Essays, Moral, Political and Literary*). He also examined how commercial development worked in tandem with moral and intellectual progress ("Of Refinement in the Arts," 1752, *Essays*). Adam Smith delineated how a framework of law and justice could permit individuals, blessed with a desire for improvement, to generate a commercial order of generalized prosperity. Smith developed, thereby, a systematic approach to trade and production that yielded insights into the coordinating movement of prices.

Significant Thinkers

David Hume (1711–1776). In his monumental work, *A Treatise of Human Nature,* Hume aimed to construct a science of human nature: The very foundation and starting point of any inquiry lies in the human subject. Drawing inspiration from Newton, Hume hoped to introduce experiment into his study of human nature, as the subtitle of his work makes clear: *Being an Attempt to Introduce the Experimental Method of Reasoning Into Moral Subjects.* Some of the experiments require us to probe our understanding or conceptions; others rely on observation or comparison, or point to historical examples. Hume applied to the operations of our minds and to society the principle of causality invoked in explanations of natural phenomena. A principle of psychological association explained the various connections and relations that unite the ideas of mind. When he turned to politics and morals, Hume employed history as a source of experiment, along with observation and comparison. On any review of history, Hume contended, one will find general relations of motive and cause that provide a basis for understanding and prediction. The basic rules of justice, he suggested, arise over time as implicit social conventions. Moreover, moral judgment rests on a mode of social communication referred to as "sympathy": the means by which the feelings of one person are communicated via inference and identification to a spectator.

Adam Ferguson (1723–1816). In *An Essay on the History of Civil Society* (1767), Ferguson offered a conjectural history of society that is less a history of particular events—the founding of nations, the outbreak of wars, and so on—than an explanation of the movement of society from primitive to commercial stages. Ferguson employed a comparative method, drawing from travelers' reports and ancient texts, to delineate a theory of human beings as cleaving to society yet relishing conflict and steadily seeking improvement. According to Ferguson, the advances of the human being are slow, piecemeal accumulations that generate rules of property and law, as well as manners and technological achievements. Unlike Hume, who believed that commerce and moral improvement worked in tandem, Ferguson worried that individualism may weaken social bonds. Moreover, the rise of luxury within commercial societies may dissipate moral vigor, thereby encouraging corruption and weakness and an inclination to serve private pleasure over public good. Ferguson is sometimes regarded as one of the first sociological thinkers. His accounts of the role of conflict as well as his emphasis on the social and moral effects of modern commerce resonated into the 19th century (influencing, e.g., G. W. F. Hegel and Karl Marx) and beyond.

Adam Smith (1723–1790). For Smith, human inquiry, including a study of society, attempts to imaginatively construct the systematic links between disconnected phenomena. Smith had hoped to develop a systematic account of society, in terms of morals, economics, and jurisprudence. However, he completed but two of these projects: *The Theory of Moral Sentiments* (1759) and *An Inquiry Into the Nature and Causes of the Wealth of Nations* (1776). In his theory of morals, Smith sought to show how a human being with proclivities to self-interest might nonetheless cooperate benevolently with others. Via a social and psychological explanation, he based moral judgment on an agreement (sympathy) in sentiment between spectator and agent, an agreement that comes about by an imaginative act in which a person puts himself in the circumstances of the other. In the *Wealth of Nations*, Smith argued against mercantilism and constraints on trade, raised questions about the economic value of imperialism and colonialism, and delineated how a system of natural liberty will increase productivity and wealth, thereby benefitting the working population. Smith relied on a notion of self-interest (less a description than an analytical hypothesis) and the desire to better our condition to reveal how trade and markets could be self-coordinating and function without a guiding hand. Although his labor theory of value proved influential into the 19th century, it was fraught with complexities. On the more positive side of the ledger, Smith's enduring legacy rests on his illumination of how the endeavors of individuals, situated within an impartial legal framework ("natural liberty"), could increase productivity and wealth without a legislating hand to guide their efforts.

Along with these significant thinkers, numerous others contributed to the understanding of society, including Henry Home, Lord Kames (1696–1782) on law (*Historical Law Tracts,* 1758; *Sketches of the History of Man,* 1774); James Burnett, Lord Monboddo (1714–1799) on the origins of language (*Of the Origin and Progress of Language,* 1773–1792); and John Millar (1735–1801) on relations of authority (*The Origin of the Distinction of Ranks,* 1771) and the evolutionary development of law (*An Historical View of the English Government,* 1787).

Concluding Remarks

None of the thinkers described above worked within established disciplines such as we know them today. Nor need we understand these inquiries as first installments in some continuous development of specific social science disciplines. However, it remains true that the Scottish inquiries proved influential to the later disciplines of anthropology, sociology, psychology, and economics. What we find in these Enlightenment thinkers is the appeal to a uniformity of human nature, the idea of causal law as relevant to the social as well as natural realm, the importance of a self-coordinating model of economic production, the idea of unintended yet complex social outcomes, and the use of history as a subject in its own right and as the basic framework for understanding the present.

Eugene Heath

See also Enlightenment, Critique of; Hobbes's Philosophical Method: Nature–Man–Society; Homo Economicus; Invisible Hand Explanations; Modernity; Montesquieu and the Rise of Social Science;

Naturalism in Social Science; Newtonianism in Adam Smith's Social Science; Spontaneous Order

Further Readings

Barfoot, M. (1990). Hume and the culture of science in the early eighteenth century. In M. A. Stewart (Ed.), *Studies in the philosophy of the Scottish Enlightenment* (pp. 151–190). Oxford, England: Clarendon Press.

Berry, C. (1997). *Social theory of the Scottish Enlightenment*. Edinburgh, Scotland: Edinburgh University Press.

Bryson, G. (1945). *Man and society: The Scottish inquiry of the eighteenth century*. Princeton, NJ: Princeton University Press.

Campbell, R. H., & Skinner, A. S. (Eds.). (1982). *The origins and nature of the Scottish Enlightenment*. Edinburgh, Scotland: John Donald.

Foucault, M. (1970). *The order of things: An archaeology of the human sciences*. New York, NY: Random House.

Fox, C., Porter, R., & Wokler, R. (Eds.). (1995). *Inventing human science: Eighteenth-century domains*. Berkeley: University of California.

Gordon, S. (1991). The Scottish Enlightenment of the eighteenth century. In *The history and philosophy of social science* (chap. 7, pp. 111–147). London, England: Routledge.

Hamowy, R. (1987). *The Scottish Enlightenment and the theory of spontaneous order*. Carbondale: Southern Illinois University Press.

Hill, L. (2001). Eighteenth-century anticipations of the sociology of conflict: The case of Adam Ferguson. *Journal of the History of Ideas, 62,* 281–299.

Hont, I., & Ignatieff, M. (Eds.). (1983). *Wealth and virtue: The shaping of political economy in the Scottish Enlightenment*. Cambridge, England: Cambridge University Press.

Jones, P. (Ed.). (1988). *Philosophy and science in the Scottish Enlightenment*. Edinburgh, Scotland: John Donald.

Meek, R. (1776). *Social science and the ignoble savage*. Cambridge, England: Cambridge University Press.

Olson, R. G. (1993). *The emergence of the social sciences, 1642–1792*. New York, NY: Twayne.

Phillips, M. S. (2000). *Society and sentiment: Genres of historical writing in Britain, 1740–1820*. Princeton, NJ: Princeton University Press.

Skinner, A. S. (1982). A Scottish contribution to Marxist sociology? In I. Bradley & M. Howard (Eds.), *Classical and Marxian political economy: Essays in honor of Ronald L. Meek* (pp. 79–114). New York, NY: St. Martin's Press.

Searle and the Construction of Social Reality

In *The Construction of Social Reality,* published in 1995, John Searle proposed a theory as to how the social world fits naturally into a physical world. Searle's theory is grounded in one core idea that unifies all of his philosophical thought: There is one world constituted by physical matter, and whatever exists in this world must be physical. It follows that a proper explanation of any phenomenon of this world must not violate the physical laws of current scientific paradigms. In light of the underlying presupposition of the physical unity of the universe, a tension immediately arises when we look around at the world in which we actually live. What we see is a world of objects that are not easily explained in terms of scientific reductions. For example, our daily lives consist of social interactions that depend on this mysterious production of sounds we know as language—with objects whose very existence depends on this language we produce. Searle's favorite example is that of money. Money might be realized in the paper or metal that you have in your pocket, but the possibility of its use as money resides in something beyond the mere physical stuff of which it is made and its linguistic representation.

Searle therefore needed to build a theory that could establish a set of necessary features for phenomena to exist as social objects, and then explain what relation these features have to the physical world. In the same way that science can tell us how the physical world is constructed from elementary particles, Searle intended to tell us how the social world in all its manifestations is constructed by human beings, by virtue of physical, biological features. Ultimately, his theory of social construction, if successful, would show higher-level social life to be unified with our lower-level foundational world of physics.

Background: Shaping Searle's Thought on Society

The path to Searle's social construction theory is historical. Beginning with his first book, *Speech Acts,* published in 1969, and continuing throughout his publications, including his latest book, *Making the Social World: The Structure of Human Civilization*

(2010), most of the elements that would eventually go into making his fully developed social theory have been gradually elaborated and put together.

Speech Acts

In *Speech Acts,* the foundational elements for his theory of social reality are first laid. In this extended essay on the philosophy of language, Searle introduced three key notions that would serve as the building blocks of human institutional reality. These are (1) the particular form of speech act that is a *declaration;* (2) the distinction between different kinds of *rules* that shape human activity, regulative and constitutive; and (3) the logical *formula* "X counts as Y in context C," which describes all human institutional facts.

1. Declarations are a particular type of speech act that enable the distinctly human capacity to both represent and create facts at the same time. Thus, by means of sincerely and explicitly stating that such and such is now the case, a new state of affairs is brought into existence by the simple act of representing it as being the case. Adjourning meetings, marrying people, christening ships or children are all acts of declarations. Although declarations can be performed by a single individual—for example, a child could say alone in her room, "From now on, this doll will hereby be called 'Ida'"—most declarations require recognized authority, incorporating more than one person as speaker in order to be successful. The declaration would prove pivotal in Searle's theory of social reality.

2. The second crucial notion for Searle's social construction theory in *Speech Acts* is about the kind of rules used for social construction. Rules are of two different types depending on the function of the rule itself. According to Searle, the function of rules that guide or direct already established human behavior is *regulative,* whereas the function of rules that create new forms of human behavior is *constitutive.* For example, driving a car down the street does not require a rule in order to engage in the activity of driving. But driving the car on one side of the road or other in order to abide by an established convention or directive is an activity that is rule directed. On the other hand, playing a game such as chess is only made possible by virtue of a set of rules that creates the game in question. Our daily life is a web of these rules, from nonsocial acts such as driving down a country road to social acts of eating together, walking down a crowded sidewalk, or getting married. Even our very births and deaths are structured by rules that shape our first appearance in this life to our very last breath in it. For example, a baby born in most modern societies must have a name by law and a birth certificate. A dying person must have a death certificate. Humans have a unique form of behavior differing from other kinds of animals in that they not only can have habitual, conventional forms of behavior that become codified, they can create new forms of activities for themselves by virtue of the very act of codification. And the causal effects of this codification on our daily life are just a powerful as the causal effects of natural events such as a thunderstorm, a drought, or a forest fire on wild animals living in nature.

3. It is also in *Speech Acts* that Searle introduces the underlying logical formula for human institutions, "X counts as Y in context C." At this point in his writing, the formula is presented as the logical structure of language. It is used as an extension of his discussion on rule distinction to illustrate how the very tool that humans use—language—is itself rule constructed. As he states, language is constructed by constitutive rules, in the form of "X counts as Y in context C," whereby making a certain string of sounds X counts as a meaningful utterance Y in context C, a given language. Searle would later employ this formula in his social theory as the logical form underlying human institutions. He would also explain in *Making the Social World* the apparent circularity that his analysis of language creates: language, which itself requires rules, creates the very rules it requires. The logical formula of "X counts as Y in context C" is without a doubt the iconic characterization of Searle's theory of social construction. As he will show, the logical formula encodes the rich cognitive processes that are required to construct social and institutional life.

Intentionality: Mind, Language, and Actions

The bridge for Searle between the physical level of the natural world and the higher-level reality of human institutions is *language.* After *Speech Acts,* Searle turned his concerns to the fundamental cognitive features essential for language. To this

end, Searle published his philosophical theory of mind in *Intentionality: An Essay in the Philosophy of Mind* (1983). There are several points in this book that would play a role in Searle's theory of social reality. First, Searle presents a theory of mind in which consciousness is an ontologically irreducible feature of the world. Second, this ontologically irreducible consciousness has causal power. These two points are vital to Searle's social theory because the single most important factor in building human reality for Searle is human collective consciousness with representational content that has causal power. Searle calls this power "intentional causation." On Searle's account, human reality is not only ontologically subjective, it exists in a kind of subjectivity that can have ontologically objective effects.

Another feature of *Intentionality* instrumental to Searle's social theory is his theory of action. Searle's social construction analyzes human reality as a blend of two levels, action and language. Part and parcel of human social and institutional reality is the ability of humans to act collectively, along with another capacity to use language in order to create social objects. A theory of action provides Searle with the link to explain how the two, language and action, work together to construct social reality. Searle's theory of action analyzes intentional human behavior in terms of three logical components: a prior intention, an intention-in-action, and a bodily movement. Prior intentions and intentions-in-action both have contents that are causally self-reflexive. Causal self-reflexivity allows intentional contents to cause their own conditions of satisfaction. Thus, if a prior intention is successful, it will cause an intention-in-action by virtue of causally realizing its contents. Similarly, the intention-in-action enables the realization of its conditions of satisfaction by means of a bodily movement. An example of this that Searle often uses is the simple action of raising one's arm. A prior intention to raise one's arm has the content "I intend to raise my arm," whereby if successful, one's arm rises, fulfilling the contents of both the prior intention and the intention-in-action.

In his analysis of action, Searle explains how our actions can be categorized in two different ways. Actions can serve as means to ends, or they can be meaningful in and of themselves. Searle calls the former by-means-of actions and the latter, by-way-of actions. An example of a by-means-of action is that of shooting a target. In order to shoot a target, a gun must be fired. This action of firing the gun is a means by which the target gets shot. The by-means-of action can consist of many auxiliary actions depending on the skill of the agent. In the case of target shooting, if a shooter is a skilled marksman, his prior intention and intention-in-action have the simple content of "I want to shoot the target," which if successful is satisfied by the simple by-means-of action of shooting the target. On the other hand, if the agent is a novice shooter, the intention of shooting the target might be broken down into several by-means-of actions, such as first aiming the gun, pulling the trigger, and then firing the gun. Now let us imagine that the marksman is in a competition and those who wish to continue in the competition must intentionally shoot the target to indicate their willingness to continue. Shooting the target then becomes a by-way-of action in that the bodily movement takes on two levels of representation: It is indeed a goal-directed action of shooting, but this action fulfills an intention that is secondary to the primary intention of signifying the intentional content: "I wish to remain in the competition."

The distinction of kinds of action would become useful to Searle later in his social theory in order to explain how individuals can collectively perform single goal-directed actions by means of individual performances. Furthermore, it would also be useful in explaining how performances acquire a double level of collective intentionality, as in the case of declarations whereby individuals collectively authorize the imposition of status functions on physical objects. Searle's analysis of action is inextricably tied to his analysis of language and society.

Collective Intentionality

In "Collective Intentions and Actions" (1990), Searle began constructing his social theory by analyzing cooperative action as the basis of all social behavior. It was in this article that Searle introduced a particular kind of psychological state, *collective intentionality,* as required for any kind of cooperative behavior. Although the notion of a psychological state consisting of shared content was not new at the time this article was written, Searle's notion of it was different in several ways from traditional notions. Rather than attempting to give an ontological definition of content or shared content, Searle establishes the essential features for psychological states to be collectively held: First, the collective psychological state must have a plural first-person agent, a "we"

rather than an "I," and, second, the collective psychological state must be directed toward a common goal or target.

Furthermore, because first-person-plural agency, along with a common goal, defines a collective intentional state and distinguishes it from an individual intentional state, the state cannot logically be decomposed into a cluster of I-intentional states mutually held among a collective. That is to say, a we-thought is a primitive, irreducible psychological state. For example, the psychological states required for two people to take a walk together is different from the psychological states held by two people who just happen to find themselves in pace with one another on the sidewalk. In the first case, the two people have intentionally created not merely an action of coordinated behavior but a coordinated behavior directed toward a single goal—taking a walk. In the second case, two individuals have different goals that happen to coordinate their behavior. In the first case, the cooperative walkers might think to themselves, "We are taking a walk." In the second case, the happenstance walkers do not think to themselves, "We are taking a walk," unless in fact, they decide to do so, at which point the goal of their intentional movement becomes common and their intentionality has shifted from I-intend to we-intend.

Searle allows for the possibility of error in grasping the fact of another's intentions in a collective action. He provides an example of two people pushing a car, one of whom is committed to a collective endeavor while the other is only pretending to push. In this case, the individual committed to the collective endeavor thinks, "We are pushing the car," but he is mistaken as to the intentionality of the other person. At first glance, this example seems to weaken Searle's claim for the irreducibility of we-intentionality, although it is not clear exactly where that weakness lies. Searle never makes the claim that a cooperative endeavor necessitates we-intentional states in all the participants. Surely in group actions, the possibility exists for individuals to simply "go along for the ride," without commitment to any goal whatsoever. The question is, what effect does the possibility of mistaking the contents of other minds have on the claim that collective intentionality exists as a primitive psychological state?

The problem is epistemic and does not affect Searle's ontological claim. Searle's ontological claim of collective intentionality has deep implications: Cooperative social behavior requires an individual to be able to conceive of himself in terms of being a plural agent, a "we" rather than an "I." Without this capacity, individuals could not engage in any cooperative social behavior whatsoever. Indeed, they could not even pretend to cooperate. In the case of the pretend car pusher, he presents himself as a cooperator insincerely in the same way an individual engaged in lying presents himself as telling the truth. Knowing how to cooperate first involves understanding what a plural agent is, along with understanding that plural agents have common goals. And this notion of agency is logically distinct from that of actions involving multiple individual agents whose personal goals converge, intentionally or unintentionally. Having argued for the necessity of a psychological state for cooperation, that is, collective intentionality, Searle then had to explain how collective intentionality could exist in individual minds yet motivate individual bodies to act. In other words, Searle needed to defend methodological individualism, inherent in his theory of mind, in light of collective intentionality.

Searle's theory of action already contained the key notion for maintaining both methodological individualism and collective intentionality. Specifically, the notion of by-means-of actions could be employed to explain how individual actions contribute to the greater goal of a collective action. Within a cooperative endeavor, each individual action is directed toward a goal that serves a larger purpose of fulfilling the overarching goal of the collective act. If the collective goal is to achieve x, each individual does his or her part, y, z, as a by-means-of action toward achieving the ultimate goal, x. As an illustration, Searle uses the example of making a sauce béarnaise, in which one person stirs while the other pours. Their overarching goal is to make the sauce, x, by-means-of their individual actions, y, z. The individual actions can occur outside the collective endeavor—for example, a person can make a sauce by pouring and stirring on his or her own—but the conditions of satisfaction of those actions are quite different from those within the collective endeavor in that the actions performed are not directed to a collective goal held by a plural agent.

As in the case of individual actions, collective intentionality rises to the level of background skill of the cooperators. For example, if the Emerson String Quartet decides to play Mozart's String Quartet No. 1, the performers play their respective parts without requiring multiple representations of

by-means-of actions of the intricacies of the music. The accomplished performance of the cello is a by-means-of action as one action in the same way as the accomplished performances of the viola and the violins are. On the other hand, the beginning grammar school quartet playing the same piece must contribute to the collective goal in bits and pieces, fits and starts, by means of multiple by-means-of actions. In individual actions, by-means-of actions are construed as means to personal ends. Applied to collective actions, by-means-of actions are construed as means to a collective goal that belongs to a plural agent, a "we."

Searle's Theory Shaped

By 1992, Searle had all the pieces together to form his theory of social reality and did so in *The Construction of Social Reality*. In this book, Searle questions the metaphysics of our everyday human life. Why is money, money? What makes an institution essentially the thing that it is? In other words, Searle questioned the ontology of our social objects: What are these things that make up our human world? Where did they come from?

Searle had long before noted in *Minds, Brains and Science* (1984) that there is a disjunct between the social world, the world of social objects, and the physical world, the world of physical phenomena. The disjunction was particularly obvious in the social sciences and the physical sciences. The physical sciences had achieved great success as a field by virtue of physical laws that could predict and explain physical phenomena. Although not for lack of trying, the social sciences never got so far as establishing any strict deterministic laws. Searle claimed that the reason for this gulf lay in the metaphysics of the objects under investigation. Whereas physical objects are constituted by matter, mental objects are essentially constituted by people's attitudes. All social objects, from tables and chairs to families, banking institutions, and national borders are what they are because of people's beliefs. Searle concluded that intentionality is the essential feature of the social world and, furthermore, although it is an extension of the physical world, it cannot be ontologically reducible to it.

In light of this, Searle needed to explain how in fact intentionality could create social objects from physical phenomena. To that end, he had what he called "the building blocks" of social reality already in place. There are three main building blocks: (1) social objects exist by virtue of collective attitudes, collective intentionality; (2) their realization comes about because individuals collectively impose a new metaphysical status on them, thereby giving them a status function, endowing them with new social roles; and (3) this collective endeavor is codifiable in principle. That is, this collective imposition of status functions on objects is by its very nature a constitutive rule; thus, it is not a coincidence that this collective endeavor has the logical structure "X counts as Y in context C," the structure Searle introduced in *Speech Acts* to explain the nature of a rule. In the case of social construction, the X of the formula is an object that is collectively chosen to undergo status function imposition. It becomes a Y by virtue of receiving its new status function, endowing it with a new social role in context C—a cultural context surrounding it.

The logical formula is the defining structure underlying Searle's theory of social construction. The simple formula proved to be powerful in several ways. One way is that it could be used to define a set of social objects, namely, institutions that differed essentially from merely physical objects such as tables and chairs by virtue of the types of functions they could serve.

Searle claims that functions are not innate to objects but are imposed on them. He stratifies the kinds of functions that humans can create into two basic types. (1) Humans can use objects instrumentally by virtue of their physical properties. This function is called *agentive* and can be imposed on an object by one or many users. If two people decide that a big rock serves well as a nutcracker, this rock can be a social object by virtue of its instrumental use, its physical structure. (2) A second type of function is *nonagentive*. Nonagentive functions perform not by virtue of physical structure alone but by virtue of a collective imposition of a status on it. Objects that receive this new status gain a particular new force, a *deontic power*. Deontic power endows social enablements of rights, responsibilities, and obligations to an object in the context of a particular culture. Thus, institutional objects, ranging from people to brute physical objects, such as pieces of paper used as money, driver's licenses, and birth certificates, are a special kind of social object. They are institutional objects that have important

social functions in a society. Indeed, they stabilize social relations within society by virtue of collective acceptance. They also have the benefit of being accessible to the collective consciousness at any time because of their codifiability. Institutions that lose their power are those that lose collective acceptance or fade out of collective memory. For example, the 19th-century institution of calling cards for home visits is no longer alive in the 21st century. Most people could not recall why they were needed anyway. Their function has faded from collective memory by virtue of past collective neglect or rejection.

A success of the logical formula is that it forms a natural connection between the brute physical world and the world of intentionality. This natural connection forms a web of relations between objects in interesting ways. Searle claims that the formula can be iterated upward, such that the basic X that has a physical structure receives a status function Y, creating a new object, which then can serve as a an X, open to imposition of another status function, Y, and so on. For example, a female X can become a wife Y in some cultural context C. That very wife X can become a mother Y in the context of the family C, and so on. In this way, the logical formula can illustrate how one object can bear many social roles. But the number of social roles that an object can bear is not limitless and certainly not predictable. In any upward construction that is meaningful—that is, the variables of X, Y, and C have semantic value—the variables are constrained by their contexts. That is to say, not any X that satisfies the generic requirements for a social function can indeed be a Y term in given contexts. In the female–wife–mother example, the iteration of X into Y only works given the cultural expectations that females are the sole candidates to be wives and mothers. Beyond the trio of female–wife–mother, the iteration of social roles can continue for a given X, but it is completely dependent on cultural expectations rather than on the properties of the X term itself. In some societies the social statuses (the Y function) of mother–grandmother is the upper limit of iteration for a female. On the other hand, in some societies the social roles of motherhood and marriage have changed their institutions by expanding the domain of who can satisfy the function of wife or mother. Gay marriage is a good example in which the function of marriage and parenting has changed how the iteration remains logically the same yet the variables differ radically: An X that is a candidate for the function of marriage is not gender dependent in gay marriage, creating the possibility of two discrete Xs with the same status function of Y (wife or husband) in the same marriage.

Searle's Recent Elaboration of His Theory

In 2010, Searle published *Making the Social World: The Structure of Human Civilization*. In this book, he elaborates on aspects of the theory that he introduced in *The Construction of Social Reality*. There are two discussions that are pivotal to the construction of the social theory itself. These are (1) the reconsideration of the psychological attitude of the collective required for participation in and maintenance of social structures and (2) categorizing the logical formula as an implicit- or explicit-declaration speech act.

In the initial formulation of the logical formula, Searle claimed that for social statuses to be imposed, the collective had to have some form of collective acceptance toward the fact of status function imposition on some object. Logically, collective acceptance functioned as an operator over the entire formula, as in "We accept that X counts as Y in context C." An objection to the notion of collective acceptance was that it made social-institutional life sound agreeable to the collective at large. Institutions such as slavery or institutional facts such as war appeared to be a kind of happy collective cooperative activity. But this was not Searle's intended meaning of *collective acceptance*. Rather, he intended collective acceptance to reflect a notion more in line with collective recognition when he created his social theory. In *Making the Social World,* he reconsidered the notion of collective participation in social contexts in which not all participants are cooperators or are agreeable to the social statuses that they find they and others bear. In this later discussion, Searle maintained that cooperation required the psychologically primitive state of collective intentionality but that one could participate in a collective without active cooperation by means of another psychological state, that of *collective recognition*.

Collective recognition is not a psychologically primitive state. It is a state in which one recognizes that a state of affairs is the case and that others also recognize this fact, along with mutual belief about

each other's states of recognition. For example, Searle again uses the case of money. The medium used for money is only successful because of collective recognition. If individuals begin to realize that the medium they use for money is not collectively recognized, or that collective recognition is beginning to wane, the fact of this medium as money also begins to weaken. But the very act of using money is a cooperative behavior requiring both collective acceptance and collective recognition. Dutch guilders were accepted as money until 2002 in Europe. After 2002, a Dutch guilder could not buy a loaf of bread. The guilder lost its value because it lost its institutional status and therefore its collective acceptance, that is, its use, along with collective recognition, that is, its status as money.

Collective recognition and collective acceptance can come apart in a social setting. This is the case of individuals who find themselves in societies with which they are for one reason or another not in accord. Individuals can participate in these institutions without collectively accepting them by means of recognizing that things are the way they are and that others recognize the validity of the current state of affairs. This is a case of collective recognition without collective acceptance. There is no collective acceptance without first collective recognition. Searle thus amended his original account of social construction to allow for social participation without cooperation.

The second pivotal addition to the theory is the categorization of the logical formula as a *declaration* speech act. Although the logical formula itself, that is, "X counts as Y in C," is transparently a constitutive rule with the function of a declaration, Searle did not declare it an implicit or explicit speech act until *Making the Social World*. The motivation for this is twofold: (1) human institutions are created only by virtue of creatures who have language and (2) the deontic powers of the Y term are only realized in terms of actual human beings.

Searle's theory of social construction argues for a level of social fact that extends beyond mere cooperative behavior. Animals can act cooperatively in their various food-seeking behaviors. Searle allows for this as a kind of cooperation. But his interest is in a social reality that extends beyond mere behavior. This is an institutional reality grounded in relations of power between individuals, power relations that are not only collectively accepted and recognized but also collectively required in given contexts. This power resides in status functions that are not a matter of brute physical force but an invisible force, *deonticity*. Deonticity is a linguistic notion. There are no behavioral equivalents of the concepts of obligations, rights, and responsibilities. Because of this linguistic requirement for deonticity, only humans have social behavior that is institutional.

Thus, the logical formula "X counts as Y in C" is always a *declaration* made implicitly or explicitly by a linguistic collective with the authority to do so. In any institutional fact, there is always an implicit "We make it the case that p," whereby the collective statement both creates a fact and states a fact. The declaration not only states a new fact, this fact is one of granting powers. And all power ultimately is grounded in individuals with the capacity to exercise it.

An objection made by Barry Smith (2003) to Searle's theory of social construction in which all statuses bottom out in some physical X is the example of corporations. Corporations are what Barry Smith calls "freestanding Ys"—that is, Ys lacking a physical X. Freestanding Ys are not grounded in any physical structure. Rather, they consist merely in the words that create them, that is, they exist only by virtue of a declaration. The possibility of freestanding Ys is an interesting challenge to Searle's theory because it can be extended to other sorts of current phenomena, for example, electronic money, which might or might not have a physical grounding. This challenges the aspect of naturalism of Searle's theory of social construction. On this possibility, the naturalistic link between the physical world and human reality that Searle wishes to ensure is allegedly broken: Human reality appears only to be a matter of intentionality realized linguistically.

Searle points out in *Making the Social World* that although the declaration alone is capable of creating freestanding Ys, the power of the freestanding Ys is only enabled by people who can exercise that power. This has the effect of grounding the status function of the Ys in people's behavior. Declarations of the sort that create digital money or corporations state the powers of a set of people with respect to a given context. The status function then is a relation between powers and people who serve as the media for the realization of these powers.

Concluding Remarks

Searle's early ideas about the unity of the world come to fruition in his work on social reality. The link between the intentionality of human institutions and the brute physical world lies in the biology of humans. Human biology allows for collective intentionality and cooperation. Human biology allows for the creation of human power relations because it can grasp and understand concepts such as deonticity by virtue of the particularly human capacity of language. Language as a biological capacity allows humans to create a reality by representing it as such, in such a way as to be collectively believed, perceived, and accepted. The high-level facts of the human institutional world form a continuum from the brute physical world to the world of social behavior to the world of institutional life. The tie that binds the facts together, that unifies them into one physical world, is that of the human biological capacity for intentionality and linguistic representation.

Jennifer Hudin

See also Collective Agents; Collective Intentionality; Plural Subjects; Social Institutions; Social Ontology, Recent Theories of; Speech Acts; We-Mode, Tuomela's Theory of

Further Readings

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language.* Cambridge, England: Cambridge University Press.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind.* Cambridge, England: Cambridge University Press.

Searle, J. R. (1984). *Minds, brains and science* (chap. 5). Cambridge, MA: Harvard University Press.

Searle, J. R. (1990). Collective intentions and actions. In P. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communications* (pp. 401–416). Cambridge: MIT Press.

Searle, J. R. (1995). *The construction of social reality.* New York, NY: Free Press.

Searle, J. R. (2010). *Making the social world: The structure of human civilization.* Oxford, England: Oxford University Press.

Smith, B. (2003). John Searle: From speech acts to social reality. In B. Smith (Ed.), *John Searle* (pp. 3–33). Cambridge, England: Cambridge University Press.

Self and Essential Indexicality

What connects the topics of the self and essential indexicality is the first-person term: *I, me, mine, my,* and their cognates ("I" for short). This is because, as it is now commonly thought, (a) *I* refers to the self, (b) *I* is an indexical term, and (c) *I* is an essential term. Considerable philosophical debate has surrounded each of these ideas, and it is necessary to come at them from within that context.

Basic Claims

The idea that *I* refers to the self incorporates two claims: that *I* is a referring expression and that what it refers to is the self. In modern philosophy, René Descartes assumed both of these claims; they play an important role in the arguments of the *Second Meditation.* His immediate successors tend to agree that *I* refers, but they tend to complicate the issue of what it refers to. So John Locke drew a distinction between being a person and a human being and then floated the idea that *I* might sometimes refer to entities of the first sort and sometimes to entities of the second sort. George Berkeley narrowed the reference of *I* to the agent one is, while he conceived that agency in even more restricted terms: It is essentially the willing of ideas into existence by the imagination. David Hume denied that there is any self at all for *I* to refer to, at least if "self" means what it is usually taken to mean: something simple and continued and individual. If such an entity were to exist, Hume insisted, it would have to be observable as such when one introspects; yet no such entity is observable in this way. Immanuel Kant also doubted that *I* refers to a self, at least in the use of the term that interested him: in the phrase "I think," which, he claimed, accompanies all our representations. He even doubted that *I* is a referring expression, in this use at least. It is for him a completely empty expression, and what it represents is the mere form of consciousness.

The other two ideas, (b) and (c), arise from dissatisfaction with these disputes about (a). They must remain irresolvable until it becomes clear what precisely is at issue when one claims that *I* refers. Fairly evidently, if *I* is a referring term, it must be a singular term; each use indicates which one individual we are thinking of or speaking about on that occasion. But

this is true of a variety of terms: names, pronouns, descriptions, and demonstratives. To which category, if any, does *I* belong? Once this question was clearly and precisely formulated, two basic options quickly dominated discussion. One is that *I* belongs to the class of names. In the 19th century, John Stuart Mill thought so, and Gottlob Frege agreed, though he complicated the issue: *I* is an ordinary sort of name when we use it in a public and communicative way but a wholly special sort of name when we use it privately. The other basic option is that *I* is a descriptive term, one that is synonymous with some particular descriptive phrase. Bertrand Russell, in the early 20th century, took this view at one stage, making *I* synonymous with "the subject of the present experience." Later, Hans Reichenbach offered a language-based alternative: "the person who utters this token." Debate between these basic options quickly exposed certain features of *I,* which prevent it from belonging to either the class of names or the group of descriptions.

One set of features leads to (b), the idea that *I* is an indexical term. What prevents *I* from belonging to the class of names, at least if we focus on the paradigm case (personal proper names), is that its reference is radically context sensitive. Whereas the name "N. N." refers on each occasion of use to whichever one person was so baptized, no matter what the context, *I* may have as many different referents as it has uses. Whom it refers to depends on the context in which it is used, on who happens to be speaking, for example. This makes *I* like *here* and *now,* at least in their standard uses: The person, place, and time to which each use refers depend on the who, where, and when of that use, respectively. So it is natural to explain the feature that is common to these terms, their so-called indexicality, in the way David Kaplan does, by appeal to rules. On his view, now standard, what makes indexicals special among the varieties of referring term is that the rule associated with each indexical (he calls it its *character*) is sufficient to give its meaning and determine the reference of each use in context. These are rules like "Any use of *I* refers to whoever uses it" and (for the central cases at least) "Any use of *now* refers to whenever it is used" and "Any use of *here* refers to wherever it is used."

Another set of features leads to (c), the idea that *I* is an essential term. What makes it unlikely that *I* belongs to the group of descriptive terms is that no one has yet managed to come up with a description that is a genuine synonym for the term. One test of whether a description A is a genuine synonym for a term B is that if one knows that A is thus-and-so, one should know that B is thus-and-so. But none of the candidate descriptions passes this test. Consider Russell's option: I may know that the subject of the present experience is unhappy, without knowing that I am unhappy—if I am going on reflections of faces, for example, and do not realize that this unhappy face is mine. The reasons why *I* cannot be replaced by a synonym give it a crucial role in our thinking, as John Perry and others have pointed out. They regard *I* as an "essential" term because it has a unique role in manifesting self-reflexive self-consciousness and because such consciousness is indispensable to us. The argument for this latter claim usually centers on action—more specifically, on the relationship between such consciousness and motivation. But the evidence is precisely that to which we have appealed in discounting the description view of *I*. In the case given above, for example, I will change my expression when, but only when, I recognize that it is *I* who am looking unhappy. For any of the various candidate descriptions, I might well know that "the F is thus-and-so," without recognizing the crucial bearing of that information on me. Conversely, it makes no sense to suppose that I might know "I am thus-and-so" and fail to recognize the bearing of that information on me. And since this feat of self-conscious self-reference partly explains how and why we are regularly motivated to act as we do, *I* is regarded as an essential term.

Implications

Suppose that the uses of *I* are sufficiently determined in context by the rule that they refer on each such occasion to whomever then uses them. We may assume that in any genuine case of *I*-use, there is at least and at most one user. Then it seems to follow that every use must be guaranteed to have a referent: the user responsible for that use. This is consistent with a view that is generally held (indeed, it accounts for that view)—that it would make no sense to suppose that any use of *I* could fail to refer. There is another view, also commonly held—that it would make no sense to suppose that in using *I* to express thoughts about oneself, one would need to identify which item was being referred to. This view

is also consistent with, and explained by, the now standard view. Suppose that uses of *I* are sufficiently determined in context by the familiar rule. We may assume that in any genuine case of a person using *I* to express thoughts, that person must recognize that she is the user. Then it seems to follow that the user who knows the rule, and knows that she is the user, must know that she is the referent of the use, without having to identify herself with the item being referred to.

This is undoubtedly why the standard view is generally held—because it seems to provide a way of combining (a) to (c), arrived at after a long, historical debate, with two views that are commonly assumed to be true: that *I* is guaranteed to refer and that its use to express thoughts is independent of identification. But there are tensions within the position. The most significant concerns the relation between (a) and these two common assumptions about *I*. If *I* refers to the self, it must be a genuine singular referring expression. But if we look at the wide group of such expressions, we notice one characteristic they all have in common. Such expressions are used to single out items, and this is because their successful use represents an achievement: when the intended item is indeed identified, singled out, and referred to. Now there is no such achievement with the use of *I,* on the standard view; the intended item need not be identified by the user, and the use could not fail to refer. So it is unclear whether uses of *I* can genuinely be said to single out items, and hence whether the term does indeed share that underlying characteristic required for membership as a singular referring expression.

Current debate about *I* reflects this tension. Can what is special about the term (e.g., its guaranteed reference and identification-independence) be reconciled with what is ordinary about it (e.g., its status as a genuine singular referring term)? The range of responses reduces to four basic options. The orthodox view is optimistic: There really is a way of reconciling these apparently contradictory properties. The nonorthodox views are pessimistic and content to be so: There is no such way of reconciling these properties, and that is because there is simply nothing to reconcile. Some deny that there is really anything special about *I;* it is more secure than other terms, but it is not guaranteed to refer, and it does depend on some form of identification. This position is held by those who regard *I* as a deictic term, like the other personal pronouns (*you, he/she*). Some deny that there is really anything ordinary about *I;* it is too special to be a genuine referring expression. This position is held by some Wittgensteinians. Some deny that there is really one term at issue here; there are two types of term, and hence there is no combination problem: We simply attribute special properties to one type and ordinary properties to the other. This position is held, for example, by those who hold variations of the Fregean idea, that there is an *I* of communication (ordinary) and an *I* of soliloquy (special). These nonorthodox views will continue to disturb the peace unless and until the tensions that give rise to them are resolved.

Maximilian de Gaynesford

See also Ego; Identity, Personal (Philosophy of); Introspection (Philosophical Psychology); Language, Philosophy of; Self-Knowledge

Further Readings

Descartes, R. (1984). *The philosophical writings of Descartes* (3 vols.; Vols. 1–2: J. Cottingham, R. Stoothoff, & D. Murdoch, Trans.; Vol. 3: A. Kenny, Trans.). Cambridge, England: Cambridge University Press.

Evans, G. (1982). *The varieties of reference.* Oxford, England: Oxford University Press.

Frege, G. (1967). The thought: A logical inquiry. In P. F. Strawson (Ed.), *Philosophical logic* (pp. 17–38). Oxford, England: Oxford University Press.

Gaynesford, M. de. (2006). *I: The meaning of the first person term.* Oxford, England: Oxford University Press.

Hume, D. (1978). *A treatise of human nature* (L. A. Selby-Bigge & P. H. Nidditch, Eds.). Oxford, England: Oxford University Press.

Kant, I. (1933). *Critique of pure reason* (N. K. Smith, Trans.). London, England: Macmillan.

Kaplan, D. (1989). "Demonstratives" and "Afterthoughts." In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–614). Oxford, England: Oxford University Press.

Locke, J. (1975). *An essay concerning human understanding* (P. H. Nidditch, Ed.). Oxford, England: Oxford University Press.

Mill, J. S. (1967). *A system of logic.* London, England: Longman, Green. (Original work published 1843)

Perry, J. (1979). The problem of the essential indexical. In Q. Cassam (Ed.), *Self-knowledge* (pp. 167–183). Oxford, England: Oxford University Press.

Reichenbach, H. (1966). *Elements of symbolic logic*. London, England: Collier-Macmillan. (Original work published 1947)

Russell, B. (1912). *The problems of philosophy*. Oxford, England: Oxford University Press.

Russell, B. (1917). Knowledge by acquaintance and knowledge by description. In B. Russell (Ed.), *Mysticism and logic* (pp. 209–232). London, England: Longman, Green. (Original work published 1911)

Russell, B. (1956). On the nature of acquaintance. In B. Russell & R. C. Marsh (Eds.), *Logic and knowledge* (pp. 125–174). London, England: Allen & Unwin. (Original work published 1914)

Strawson, P. F. (1966). *The bounds of sense*. London, England: Methuen.

Self and the Social Sciences

The self lies at the center of mental life. As the pragmatist philosopher and psychologist William James noted in the *Principles of Psychology* (1890/1981):

> Every thought tends to be part of a personal consciousness. . . . It seems as if the elementary psychic fact were not thought or this thought or that thought but my thought, every thought being owned. . . . On these terms the personal self rather than the thought might be treated as the immediate datum in psychology. The universal conscious fact is not "feelings and thoughts exist" but "I think and I feel." (p. 221)

Brief Historical Review

James's monumental work on the self (a chapter that covered more than 100 pages!) set the stage for much subsequent work on the topic in the social sciences. However, events taking place in the early 1900s put into play a number of assumptions about human behavior that resulted in wide-scale cessation of self-related research in academic psychology. Specifically, the behaviorist movement, primarily associated with the work of John Watson in the United States and Ivan Pavlov in Russia, came to dominate the research agenda for academicians in the psychological sciences. Behaviorists treated the organism as a "black box" that connected stimuli with responses but whose internal workings could safely be ignored. As a result, the self, clearly a part of the content of the black box, did not fit this theoretical worldview.

Fortunately, the decision to ban the workings of the black box from academic psychology did not transfer to disciplines in the social sciences outside the influence of behaviorism. For example, research on the self by sociologists was largely unabated. One prominent movement involving the self is exemplified by the work of the *symbolic integrationists*, such as Charles Cooley and George Herbert Mead, who, taking their lead from James's speculations about the social aspects of self, argued that the self was, at least in part, a social creation. Thus, Cooley introduced the metaphor of the "looking-glass self" to underscore the idea that we come to know who and what we are by seeing ourselves as reflected in the eyes of important others with whom we interact.

Another sociological enterprise focused on the self and its relation to the individual's temporal orientation. Such work (largely in the 1950s and 1960s) was characterized by a concern with (a) the factors influencing a person to temporally orient his or her self toward the past, present, or future and (b) the consequences of a temporal perspective for one's social-psychological function and mental health.

The self also was alive and well in the domains of clinical psychology and psychiatry. Starting with the far-reaching impact of Freudian thought on clinical practice around the start of the 20th century, the self (or ego/identity) assumed a central role in therapeutic thought and practice. Freudian theory and its implementation subsequently were supplemented and expanded both by Neo-Freudian perspectives on the self (e.g., by psychologists such as Alfred Adler, Karen Horney, and Otto Rank) and by competing analytic views (e.g., those of Eric Erickson, C. G. Jung, and Harry Stack Sullivan). But the ego remained a central concern. Medical schools soon became the new home, within psychology, for the study of self.

At the close of the 1950s, all this began to change as academic psychologists slowly freed themselves from the overly restrictive theoretical commitments of behaviorism. Reflecting this change, the number of research papers exploring self-related processes in academic psychology since the early 1950s has been proliferating at a staggering pace. Commenting in 1971 on the explosion of empirical articles on the self appearing in the 20 years since the behaviorist movement in American had gradually given way

to the traditional appreciation of the importance of inferring mental structures from the behavior of persons, Kenneth Gergen recorded an astonishing 2,000 studies published on the self during the previous two decades. Approximately 17 years later, John Kihlstrom and his colleagues identified dozens of "hot" research areas in which the word *self* served as a prefix (thus occupying the central position with regard to the topics addressed)—for example, self-awareness, self-concept, self-control, self-identity, self-verification, self-perception, self-handicapping, self-reference, self-regulation, self-schema, self-image, and self-stereotypes.

The Problem of Self in the Social Sciences

While it is clear that the self now has become central to the social sciences in the manner James understood it should be, it would be wrong to conclude that the intense research it receives has done much to clarify exactly what we refer to when we study the vast variety of self-hyphenated topics that populate research journals in a wide variety of domains (e.g., sociology, anthropology, and clinical, developmental, cognitive, personality, and social psychology). What is the self that regulates, references, handicaps, serves as a concept, and enables awareness?

One source of difficulty is that scientists studying the self often do not differentiate between aspects of and the ways in which they use the word *self*—that is, as an object of scientific analysis versus as first-person experience. Bertrand Russell's distinction between knowledge by acquaintance and knowledge by description provides a way to make sense of this definitional confusion. Russell proposed that we have knowledge by acquaintance when we know something by direct personal contact (sensory or introspective) and that we exhibit that knowledge by using appropriately referential terms when we communicate with others. With respect to the self, this is seen in the ease with which I talk about the self as well as understand talk about the self by others.

However, when we attempt to make explicit what it is we refer to by the word *self*—when asked to describe what the word *self* means—problems quickly arise. What exactly is the "self" that serves as the object of this diverse set of predicates? What is it that is being verified, conceptualized, controlled, esteemed, deceived, referenced, regulated, depleted, and handicapped? Although this question has led to models of the self dating back more than 100 years, it is difficult to find a coherent and convincing descriptive account of the self, per se. Rather, most papers typically describe the neurological and cognitive mechanisms that appear causally responsible for knowledge available to the self. In short, it has proven notoriously difficult to provide a set of propositions capable of transforming our acquired knowledge into a satisfying description of what a self is.

Some have argued that this really is not an issue at all. Researchers know that the object of the predicate in the relation X (self)–Y (concept, depletion, image, regulation, reference, etc.) is really a "place holder" for the modifier (Y), which is the real focus of examination. However, if that is the case, then we run into uncertainty vis-à-vis which aspect of the X–Y relation is doing the causal work. If the former (X), then in what sense is a self causally efficacious such that it can, for example, be depleted? Does it lose air, phlogiston, ether, spreading activation? If Y, then how does depletion casually work on a construct that would appear to lack substance under the physicalist principle of causal closure?

The point is that if one wishes to discuss the self and not simply see it as a slot to be filled in at some future date, then theory and research that invoke it as a causally relevant factor in psychological mechanism need to give it the respect and attention it requires.

The Neural Self of Science and the Subjective Self of First-Person Phenomenology

This section turns to what can and cannot be asserted with reasonable, empirically warranted confidence about the self. Specifically, first we discuss what Stanley Klein calls the *epistemological self*—the behavioral, affective, cognitive, and neural systems assumed to be causally responsible (at least in part) for providing the subjective self with knowledge of who and what it is. The discussion then points to an apparent incompatibility between treating the self as both the content of experience (i.e., an object) and the agent of experience (i.e., a subject). We will refer to the latter aspect of the self as the *ontological self*.

Considerable progress has been made in describing the cognitive and neurological bases of the epistemological self. This is because, unlike ontological questions, the epistemological self—that

is, the neurological bases of self-knowledge—is empirically testable and thus amenable to scientific analysis. Specifically, empirical studies of the epistemological self suggest that the singular, unified self of everyday experience actually is informed by a number of different, functionally isolable neurocognitive systems. These include, but are not limited to, the following:

1. Episodic memories of one's life events
2. Semantic memory-based summary representations of one's personality traits
3. Semantic memory-based knowledge of the facts about one's life
4. *The experience of continuity through time:* The connection between the "I" experienced now and the "I" experienced at previous points (as well as later points) in one's life—episodic memory is known to contribute heavily to this ability
5. *A sense of personal agency and ownership:* The belief—or experience—that "I" (agency) am the cause of "my own" (ownership) thoughts and actions
6. *The ability to self-reflect:* To form metarepresentations where the agent is the self and make inferences on the basis of those representations
7. *The physical self:* The ability to represent and recognize (e.g., in mirrors, photographs) one's body
8. *The emotional self:* The ability to experience and produce emotional states (both transient and dispositional) that provide value, affective valence, and evaluative direction to our actions and reasoning

Although the environment (social, cultural, and physical), and its reciprocal influence on the person, is not included in the above, this is not to dismiss its obvious relevance for what it means to have knowledge of one's self. However, the claims about the constituents of the epistemological self are primarily placed at the level of neural architecture. Accordingly, social and situational self-knowledge is folded into, and thus contained within, the neural machinery provided by systems of memory. This may be a vast oversimplification (as suggested by consideration of mirror neurons, theory-of-mind processes, and other neural structures that appear unique to social cognition rather than cognition taken more generally). It can be acknowledged that the selection criteria may be overly restrictive for some tastes.

Although the above sources (1–8) each contribute to the experience of self as a subjective singularity, taken individually they are functionally independent. That is, while in normal individuals, sources of self-knowledge work together to help create our sense of self as a subjective unity, taken separately none of these systems are either logically or empirically necessary to maintain the experience of the self as a singular, subjective point of view.

By contrast, the ontological self, the self of personal experience, is too poorly understood to bear the definitional adequacy required of the terms of a causal relation between self and memory. Not surprisingly, many researchers (intentionally or otherwise) sidestep this difficulty, relying on their readers' familiarity with the term *self* (i.e., the self of subjective experience), derived from years of knowledge by acquaintance, to confer a sense of confidence that he or she knows to what it is the author refers. But the basic problem remains—we are unclear what it is we are referring to when we apply the label *self* (nor is the term open to being grasped and thus labeled via scientific objectification). This is a serious problem.

Compounding this difficulty, researchers often fail to appreciate that the self as subjectivity—the ontological self—is *not* the object of their experimental tasks. Indeed, it logically cannot be the object of their studies. Objectivity is based on the assumption that an act or object exists independently of any individual's awareness of it; that is, it is something "other" than the self. When objectivity is the stance adopted by the self to study itself, the self must, of necessity, be directed toward what is not the self but rather some "other" that serves as the self's object. To study myself as an object, I must transform myself into an "other," into a "not-self." Thus, the self is not, and cannot, be an object for itself and still maintain its subjectivity. Paradoxically, we achieve "objectivity" of the ontological self only at the cost of losing awareness of our self as a subjective center.

The ontological self thus would seem a poor candidate for "current" scientific exploration—an enterprise predicated on understanding objects and their relations. Science is the world of publicly observable and physically measurable objects and events. Since nothing can be an object for the self unless it is "other" to the self, it follows that the self cannot objectively apprehend itself as itself. For the subjective self to become a part of the scientific world, it

would have to forfeit its subjectivity. Scientific analysis therefore has the unintended consequence of eliminating the object under discussion—the ontological self—from the discussion.

Conclusion

When social science researchers focus on the self, what they are investigating, more often than not, is actually the multiplicity of social and neural systems assumed to present the ontological self with knowledge. A tacit assumption is that there is a substantive, objective self, which, like any object (provided proper tools were available), can be treated as "other" and, thus found, grasped, and studied scientifically. While this assumption has its merits for the study of epistemological sources of self-knowledge, we often fail to sufficiently appreciate that (a) the subjective self is not an object but an awareness, a consciousness, and as such is not privy to anyone but itself and (b) there are profoundly important differences between the self as a subjective entity (the ontological self) and the self as types of knowledge available to that subjectivity (i.e., the epistemological bases of self-knowledge). These two forms, or aspects, of "self" are contingently related but are not conceptually reducible. By conflating them, we assume that we are casting empirical light on one (the assumed, causally relevant mechanism of our research endeavors—that is, the ontological self) while experimentally exploring the other (neural-based sources of knowledge about the person we are—that is, the epistemological self).

Stanley B. Klein

See also Behaviorism in Psychological Explanation; Consciousness; Ego; Identity, Personal (Philosophy of); Personal Identity and Trauma; Identity (Social); Introspection (Philosophy of); Self and Essential Indexicality; Self-Knowledge; Social Cognition; Symbolic Interactionism; Therapy, Psychological and Philosophical Issues; Unconscious Social Behavior

Further Readings

Cooley, C. H. (1902). *Human nature and the social order.* New York, NY: Scribner.

Gergen, K. J. (1971). *The concept of self.* New York, NY: Holt, Rinehart, & Winston.

James, W. (1890). *Principles of psychology* (Vol. 1). New York, NY: Henry Holt.

Kihlstrom, J. F., Cantor, N., Albright, J. S., Chew, B. R., Klein, S. B., & Niedenthal, P. M. (1988). Information processing and the study of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 145–178). New York, NY: Academic Press.

Klein, S. B. (2012). The self and its brain. *Social Cognition, 30,* 474–518.

Mead, G. H. (1934*). Mind, self and society.* Chicago, IL: University of Chicago Press.

Russell, B. (1956). On the nature of acquaintance. In B. Russell & R. C. Marsh (Eds.), *Logic and knowledge* (pp. 125–174). London, England: Allen & Unwin. (Original work published 1914)

Sorabji, R. (2006). *Self: Ancient and modern insights about individuality, life, and death.* Chicago, IL: University of Chicago Press.

Self-Direction and Self-Ownership

The terms *self-direction* and *self-ownership* can be used in many different ways. For present purposes, it is best to take *self-direction* to refer to a possible and (putatively) morally portentous feature of persons' lives and to take *self-ownership* to refer to a central moral right that persons may be thought to possess by virtue of their respective (capacity for) self-direction and/or the value, significance, or moral import of their (capacity for) self-direction.

Self-direction is a matter of an agent's directing her life (not just her momentary behavior) on the basis of her reasons, judgments, projects, and commitments. Indeed, through self-direction, an agent in part constitutes herself as an embodiment of certain projects and commitments. However, self-direction need not take the form of hyperbolic autonomy—that is, continuous self-conscious and self-critical deliberation about the course of one's life.

Self-ownership is a matter of an individual possessing the moral *liberty-right* to dispose of herself—in other words, of the elements, attributes, and powers that constitute herself—as she sees fit and the moral *claim-right* against all others not to be prevented from exercising that liberty, that is, not to be prevented from disposing of her person as she chooses. Self-ownership is the moral liberty to dispose of oneself as one sees fit *combined with* the moral right against others that they not preclude one from doing as one chooses with one's own body and mind and one's physical and mental faculties. Many general moral rights, such as the right to determine

what one will say or the right to determine what one will ingest, are profitably seen as aspects of self-ownership. Like self-ownership itself, these rights are negative in the sense that they merely obligate others to allow one to act as one chooses with the means at one's disposal.

The thesis under consideration here is that in some way the value or significance or moral import of self-direction—or of something that is closely linked to self-direction—makes plausible the affirmation of the rights of self-ownership.

Despite claims to the contrary, there is nothing especially mysterious about the term *self-ownership*. The term is certainly no more mysterious than *self-direction*. Although the rights asserted in the name of self-ownership could be advanced without employing this locution, there is a long tradition within political thought that employs this terminology, including John Locke's *Second Treatise of Government* and Robert Nozick's *Anarchy, State and Utopia*. Moreover, the locution allows the advocate of the right to dispose of one's body and mind as one sees fit to present her position as the most thoroughgoing rejection of the idea that any person is to any degree the natural slave of others. This thoroughgoing denial of any degree of natural subordination is a corollary of the claim that no one's self-ownership imposes any positive obligations of service upon others. Finally, the employment of "self-ownership" places advocacy of persons' rights to do as they see fit with themselves within a broader program of facilitating peace and freely chosen cooperation among persons through the delineation of the moral fences that define "mine" and "thine."

If the rights of self-ownership are thought of as grounded in some yet more basic and morally portentous feature of the lives of individuals—for example, in the value, significance, or moral import of their (capacity for) self-direction—then those rights will be understood as *original* or *natural* rights. They will be understood as rights that persons have by virtue of their nature as beings capable of valuable self-direction, not merely by virtue of contract, convention, or legislative decree, or the social utility of belief in those rights. Even if human beings are not born *with* (the full range of) the rights of self-ownership, they are nevertheless, as Locke put it, born *to* self-ownership in the sense that in due course they will acquire the natural features that ground the attribution of self-ownership to them. Indeed, most defenders of self-ownership take these rights to be founded upon certain morally fertile natural features of human existence and, hence, to be original or natural rights.

In contrast to this belief in *original* or *natural* self-ownership, consider J. S. Mill's assertion in *On Liberty* that every individual is sovereign over himself, that is, over his own mind and body. Mill explicitly rejects the idea that the rights of self-sovereignty have any nonutilitarian foundation and insists instead that any ascription of moral rights must be based on a calculation of social utility—albeit a calculation that ascribes utility to individual self-direction as one of the permanent interests of mankind. However, it is precisely because Mill grounds his case for self-sovereignty on social utility that the domain of sovereign individual choice endorsed in *On Liberty* contracts steadily in the last two chapters of that work as Mill's list of socially beneficial interferences with individual choice lengthens. It seems that robust and far-reaching rights of self-ownership must be original or natural rights—in other words, rights anchored to yet more basic features of persons and, hence, rights not subject to the currents of social utility, convention, or legislation.

The affirmation of self-ownership is supposed to capture and articulate a range of our most basic intuitions about the *moral inviolability* of persons. Unless individuals have waived or forfeited their rights, they are not to be killed, struck, maimed, immobilized, or enslaved, and they are not to be coerced by the threat of such actions. They are not to be subject to such treatment even for the sake of great gains for others, for themselves, or for their righteousness. At least in their paradigmatic form, moral rights do not merely require that one be duly compensated if one is subjected to any such infringement; these rights require that one simply not be subjected to such treatment. To treat the rights of self-ownership as merely requiring that compensation be paid to those who suffer (unprovoked) infringements of them is to throw overboard the inviolability of persons that attributions of rights are supposed to safeguard.

Another important feature of robust rights of self-ownership—strongly emphasized by Nozick—is that they may not be violated even to minimize the total of rights violations. You may not kill one innocent person even if doing so will distract an aspiring evildoer who is about to kill five other innocents. Each

innocent person's self-ownership imposes upon you an obligation not to kill that innocent; with respect to each, you are subject to a "moral side constraint" against killing. You violate this moral stricture if you kill the one in order to save the five from the aspiring evildoer. But you do not violate this stricture if you abstain from killing the one. Of course, if you abstain, the evildoer proceeds to violate the self-ownership rights of each of the five. That is not a good thing; but not preventing those killings does not make you a violator of the rights of the five.

On what might robust, side-constraining rights of self-ownership rest? They cannot rest on the *agent-neutral* value of the condition that they purport to protect, that is, the lives or the well-being or the self-direction of individuals. A condition has *agent-neutral* value if and only if that value provides everyone with an equally strong reason for promoting each instance of that condition. (Different people will have different countervailing reasons of different strengths for promoting instances of competing valuable conditions.) Suppose that it is thought that the rights of individuals not to be killed rest upon the agent-neutral value of individual life or individual self-direction (both of which are eliminated by killing). On that supposition, what is really wrong with killing any individual is that this act diminishes the amount of agent-neutral value in the world. But if that is what is wrong with such killing, it is not wrong for you to kill the one to distract the evildoer from killing the five. For this act does not diminish the amount of agent-neutral value in the world. It brings about a world that is better endowed with agent-neutral value than the world would be if you eschewed killing the one. If the underlying rationale for rights were the agent-neutral value of the conditions to which persons are said to have rights, then persons will not have robust, side-constraining rights.

Robust, side-constraining rights can be a salient feature of morality only on the basis of the separate, ultimate, and, hence, *agent-relative* value of each individual's life or well-being or self-direction. A condition, say, Mary's self-direction, has *agent-relative* value if and only if the value of that condition provides the person *in whom that value would be realized,* in this case Mary, with a reason to promote it and yet may provide no other individual with a reason to promote it. Agent-relative value is decentralized value—not merely in the sense that this value occurs in many places but in the stronger sense that each person is a center of ultimate sui generis value. Only the agent-relativity of the value of persons or their lives or their self-direction explains the *irreplaceablility* of persons.

It is important to see here that the agent-relativity of the value of each person's life or well-being or self-direction serves two argumentative functions. First, it undermines agent-neutralist justifications for interfering with an individual's disposing of herself as she sees fit, namely, justifications that point to some alleged increase in the world's endowment of agent-neutral value. For instance, it undermines the assertion that your killing of the one is justified because there is more agent-neutral value in a world in which the five live than in the world in which the one lives. The agent-relativist responds that—at least in the absence of very special strong relationships of concern among some of these individuals—a world in which the five live is better for each of the five and worse for the one, period. Just as each of the five has reason to favor the outcome in which he lives, the one has reason to favor the outcome in which she lives.

However, this undermining of any agent-neutralist justification for your interference with the one does not itself establish the *wrongness* of your killing the one. It does not establish the right of the one not to be killed. The second function of the appeal to the separateness of persons—that is, to the separate importance of their lives and so on—is to provide individuals with a reason to eschew interferences with other persons and, hence, to show the wrongness of those interferences. Here is the central thought: The fact that each person has ultimately valuable ends of her own that she has reason to promote provides each *other* agent with a reason not to preclude that person from disposing of herself as she sees fit in pursuit of her ends. All other agents have reason to eschew treating this person as though she is a mere resource without purposes of her own and, hence, available for their exploitation. In more Kantian language, since persons are moral ends-in-themselves by virtue of the separate ultimate value of their lives, well-being, or self-direction, persons are not to be treated as means to the ends of others. Moral rights function to *honor* the separate importance of each person (or each person's life, well-being, or self-direction); they function to give practical effect to the moral *status* that each person has by virtue of the separate and ultimate importance of her life or well-being or self-direction.

Self-direction may enter in here in either (or both) of two ways. One may say that for each individual, value lies especially in her self-direction. It is the agent-relative value of self-direction that explains or is at the core of the agent-relative value of each person's living well. Hence, respect for the individual requires respect for her self-direction. Or one may say that for each individual, value lies especially in her living well, but a highly salient fact about human beings is that they live well through self-direction; they live well through the exercise of their ability to guide their actions by their reasons, judgments, projects, and commitments. Hence, respect for human beings as centers of value takes the form of respect for self-direction.

Eric Mack

See also Free Will, Philosophical Conceptions of; Libertarianism, Political; Normativity; Oppression; Philosophy of Politics

Further Readings

Cohen, G. A. (1995). *Self-ownership, freedom, and equality.* Cambridge, England: Cambridge University Press.

Locke, J. (1980). *Second treatise of government.* Indianapolis, IL: Hackett. (Original work published 1689)

Lomasky, L. (1987). *Persons, rights, and the moral community.* Oxford, England: Oxford University Press.

Mack, E. (2009). Individualism and libertarian rights. In J. Christman & T. Crisiano (Eds.), *Contemporary debates in political philosophy* (pp. 121–136). Oxford, England: Blackwell.

Mill, J. S. (1978). *On liberty.* Indianapolis, IL: Hackett. (Original work published 1859)

Nozick, R. (1974). *Anarchy, state and utopia.* New York, NY: Basic Books.

Rasmussen, D., & Den Uyl, D. (2005). *Norms of liberty.* University Park: Pennsylvania State University Press.

Self-Knowledge

Western philosophy has long treated self-knowledge as knowledge of *an inner self*. That use of *inner* is epistemic, designating a self knowable introspectively, that is, only by the same person. On that conception, even scientific knowledge of a person's nature would be incomplete knowledge of that person's self.

This entry describes the locus classicus of that traditional conception, plus a few prominent challenges and alternatives to that conception. Recently, some have argued for self-knowledge's including knowledge of the wider world.

Descartes: The Classic Conception

Modern philosophy began with René Descartes (1596–1650), especially his epistemological and metaphysical writings. *Discourse on the Method* and *Meditations on First Philosophy* exemplify *autobiographical* and *introspective* philosophy: Descartes tried to ascertain for himself what he knows, first questioning whether he knows anything at all. In "Meditation I," he wondered whether his beliefs ostensibly resulting from observing the physical world might be mistaken by being present within a dreaming experience. He then postulated as possible an evil demon deceiving him in all his beliefs, even those apparently due to reason.

This is where a concept of self-knowledge entered the story. In "Meditation II," Descartes reflected that with some of his beliefs it was impossible to be deceived. Being deceived involves thinking. So even when whatever one is thinking is the result of being deceived, one cannot be mistaken in thinking that one is thinking; and thinking includes existing. That was Descartes's *cogito*. Its most celebrated version is *cogito ergo sum*—"I think, therefore I am"—from the *Discourse*. Descartes claimed similarly invulnerable knowledge of his particular acts of thought—doubting, imagining, willing, and so forth. Thus, a self was described—known by that same self, immune from skeptical doubt. This is philosophy's locus classicus of putative self-knowledge.

Hume: The Self as Bundle

For Descartes, the known self was a mentally active substance. David Hume (1711–1776), in *A Treatise of Human Nature,* advocated a contrary picture, reflecting on how a self would be known. Hume allowed that there is mental activity with mental content. However (he concluded), no inner substance is knowable as the agent of that activity. We would know such a substance by introspectively discovering its accompanying the mental contents. Yet whenever we introspect to effect that discovery, we find only more mental contents. We do not discover also an inner substantial self. Hume did infer, nonetheless, that selves exist. But his was a distinct

conception: The self is a mere bundle or collection of those mental contents—not an entity, or substance, over and above those.

Kant: Two Selves

Immanuel Kant (1724–1804), in the *Critique of Pure Reason,* distinguished between the self as experienced (the *phenomenal* self) and the real self (the *noumenal* self). We can know only the former—one's self as involved in having experiences. There is no such experience—hence no knowledge—of a real self characterizable apart from one's having the experiences. Inevitably, though, we continue believing that such a self exists; we organize our thoughts as if it does. That is part of how we conceive even of experiencing a world beyond ourselves. You do not know your self as an object; you think of it as a subject.

Anscombe: Practical Self-Knowledge

Part of self-knowledge could be knowing oneself in action, including mental action. Elizabeth Anscombe (1919–2001), in *Intention,* highlighted a particular instance of that. Seemingly, knowing that one is about to eat an apple differs in kind from knowing that one's dog is about to do so. The latter knowledge is predictive; the former is also indicative of one's intending to act in that way. Anscombe called this *practical* knowledge. It would be practical self-knowledge. Her discussion has influenced many philosophers writing on action.

Semantic Externalism and Active Externalism

Recent philosophy poses another challenge to Descartes's classic conception. That challenge concerns the semantic content of thoughts. It reflects a view of meaning—semantic externalism (or "social anti-individualism and the mental" in this *Encyclopedia*)—associated especially with Hilary Putnam (1926–). Putnam imagined speakers on a twin of Earth—Twin Earth—using the term *wate*r as we do on Earth, responding to apparently the same sort of liquid—except that Twin Earth's "water" is XYZ, not H_2O. So their term's meaning differs unwittingly from that of ours.

This would imply that no one can know purely introspectively quite what his or her uses of such a term mean. To that extent, no one could know themselves by introspection. A person would have to know the surrounding world (e.g., whether this is Earth with H_2O rather than Twin Earth with XYZ) even to know his or her own mind. The former knowledge would depend on complex observation, even scientific theorizing. Self-knowledge would thus differ from Descartes's conception of such knowledge as immediate and first-personal. It would involve observing the wider world, even in order to know one's inner thoughts.

Maybe the self being known would actually be a part of the larger world. Andy Clark and David Chalmers propose an active externalism about mentality: One's mind is partly those aspects of the world helping one's functioning in cognitive respects. Diaries, calculators, even iPhones actively extend one's mind and, hence, one's self. Self-knowledge would include knowing one's iPhone.

Stephen Hetherington

See also Consciousness; Distributed Cognition and Extended-Mind Theory; Epistemology; Introspection (Philosophical Psychology); Self and Essential Indexicality

Further Readings

Anscombe, G. E. M. (1963). *Intention* (2nd ed.). Oxford, England: Basil Blackwell. (Original work published 1957)

Cassam, Q. (Ed.). (1994). *Self-knowledge.* Oxford, England: Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58,* 7–19.

Descartes, R. (1911a). *Discourse on the method* (The Philosophical Works of Descartes, Vol. 1; E. S. Haldane & G. R. T. Ross, Eds. & Trans.). Cambridge, England: Cambridge University Press. (Original work published 1637)

Descartes, R. (1911b). *Meditations on first philosophy* (The Philosophical Works of Descartes, Vol. 1; E. S. Haldane & G. R. T. Ross, Eds. & Trans.). Cambridge, England: Cambridge University Press. (Original work published 1641)

Gertler, B. (2011). *Self-knowledge.* London, England: Routledge.

Hetherington, S. (2007). *Self-knowledge: Beginning philosophy right here and now.* Peterborough, Ontario, Canada: Broadview Press.

Hume, D. (1978). *A treatise of human nature* (P. H. Nidditch, Ed.; 2nd ed.; Book I, Pt. IV, sec. VI). Oxford, England: Clarendon Press. (Original work published 1739–1740)

Kant, I. (2007). *Critique of pure reason* (N. K. Smith, Trans.; 2nd ed.; B155–B159). Basingstoke, England: Palgrave Macmillan. (Original work published 1781–1787)

Putnam, H. (1975). The meaning of "meaning." In *Philosophical papers: Vol. 2. Mind, language and reality* (pp. 215–271). Cambridge, England: Cambridge University Press.

Semantics and Pragmatics

Semantics and pragmatics are the two major branches of linguistics devoted to the study of meaning in language. This much is largely accepted. However, what constitutes the domain of semantics and that of pragmatics? Can semantics and pragmatics be distinguished in a principled way? Are they autonomous, or do they overlap with each other? How and to what extent do they interact with each other?

Semantic Minimalism Versus Contextualism

Currently, there is a heated debate between semantic minimalism and contextualism in the philosophy of language. The central thesis of semantic minimalism is that context is allowed to have only a very limited or minimal effect on the semantic content of an utterance. In addition, semantic minimalism holds that semantic content is entirely determined by syntax, semantic context-sensitivity is grammatically triggered, and it is not the job of semantic content to capture one's intuitive judgment of what a speaker says when he or she utters a sentence. Consequently, the object of study of semantics should be separated strictly from pragmatic intrusion. A number of variants of semantic minimalism can be identified. They include Emma Borg's *minimal semantics,* Hermen Cappelen and Ernest Lepore's *insensitive semantics,* and Kent Bach's *radical semantic minimalism.*

By contrast, *contextualism* endeavors to provide an account of contextual variations in semantic content in terms of a criterion of contextual best fit. According to the contextualist view, pragmatically enriched entities such as speech acts are the primary bearers of truth-conditional content. Only in the context of an utterance does a sentence express a determinate semantic content. In other words, semantics covers only part of the way toward the computation of utterance meaning, and it is pragmatic enrichment that completes the process as a whole. Currently, there are two versions of contextualism: moderate and radical contextualism. While the former acknowledges limited pragmatic influence on semantic content, the latter holds the view that pragmatic processes such as free enrichment play a central role in explaining contextual variations in semantic content. Furthermore, closely associated with radical contextualism is a position known as "truth-conditional pragmatics," namely, the thesis that various pragmatic processes influence and determine the truth-condition of an utterance.

Next, there is semantic relativism. Semantic relativism is an approach in the philosophy of language that falls largely in the semantic minimalist camp. While acknowledging that varying standards have a semantic role to play, proponents of semantic relativism reject the contextualist claim that the role in question is relevant to the determination of what is said by an utterance. Rather, the role played by varying standards is relevant to determining whether what is said is true or false. Some semantic relativists distinguish a context of use from a context of assessment and insist that epistemic standards, for example, are features of the context of assessment.

Finally, of some interest is what is called *indexicalism.* This is the position in the philosophy of language that assumes that there is a role for the speaker's meaning to play in the determination of the truth-conditional content of a sentence, but only when a slot is set up by the sentence itself to be pragmatically filled in its logical form. To this end, a range of "hidden" indexicals is posited to provide syntactic triggers for the additional context-sensitivity demanded by indexicalists, thus also referred to as *hidden indexicalism.* No top-down pragmatic influence is allowed to affect the truth-conditional content of a sentence. Interestingly enough, indexicalism is considered to be a version of moderate contextualism by semantic minimalists and a variety of semantic minimalism by contextualists. An alternative view is that, contrary to indexicalists, context sensitivity, called for by contextualists, lies in the circumstances of evaluation rather than being a truly indexical content for a sentence. This position is labeled *nonindexical contextualism.*

Drawing the Semantics/Pragmatics Distinction

The semantics/pragmatics distinction has been formulated in a variety of different ways. But of various formulations, three are particularly influential. First, it has been characterized in terms

of truth-conditional versus non–truth-conditional meaning. According to this formulation, semantics deals with truth-conditional meaning, and pragmatics has to do with non–truth-conditional meaning. This characterization of pragmatics is captured in a well-known formula proposed by Gerald Gazdar: Pragmatics = Meaning − Truth conditions ("Pragmatics equals meaning minus truth conditions"). There are, however, a number of problems at the very core of this approach to the semantics/pragmatics division. First, there are linguistic constructions that do not denote anything and therefore do not make any contribution to their truth-conditional content. Paradigmatic cases include greetings such as *Good morning,* conventional implicature-triggers such as *but,* and syntactic constructions such as imperatives. Second, the linguistically encoded meaning of a sentence does not always fully determine its truth conditions. This is illustrated by a sentence such as "It is snowing," which, according to contextualists, contains an unarticulated constituent, such as [in London]. Furthermore, there is often a pragmatic intrusion into the truth-conditional content of a sentence uttered. All this has led François Recanati to dub (part of) pragmatics *truth-conditional pragmatics.* If this is correct, one has to conclude that the truth-condition constraint cannot itself distinguish semantics from pragmatics in a principled way.

Second, the demarcation line between semantics and pragmatics has been defined in terms of conventional versus nonconventional meaning. On this view, semantics studies the conventional aspects of meaning and pragmatics, the nonconventional aspects of meaning. However, as pointed out by Kent Bach, this way of invoking the semantics/pragmatics division runs into trouble with the fact that there are linguistic expressions whose conventional meaning is closely associated with language use. A case in point is discourse indexical expressions. A major function of discourse indexical expressions such as *anyway, after all,* and *in conclusion* is to indicate that there is a relation between the utterance that contains them and some portion of the prior discourse. A further point to note is that the conventionality of a linguistic phenomenon may be a matter of more or less rather than yes or no. For example, of the three types of implicature proposed by Paul Grice, conventional implicature is the most conventional, hence the most "semantic" and the least "pragmatic"; particularized conversational implicature is the least conventional, hence the least "semantic" and the most "pragmatic," with generalized conversational implicature lying somewhere in between. In other words, the three types of implicature form a semantics–pragmatics continuum whose borderline is difficult to mark. From facts like these, one can arrive at the conclusion that there is no neat correlation between the semantics/pragmatics distinction and the conventional/nonconventional meaning distinction.

Finally, the semantics/pragmatics distinction has been equated with context independence versus context dependence. According to this formulation, if a linguistic phenomenon is invariant with respect to context, then it is the concern of semantics. By contrast, if a linguistic phenomenon is sensitive to context, then it is a topic of pragmatics. This characterization of the semantics/pragmatics distinction, however, rests on a mistaken assumption that context has no role to play in semantics. According to Bach, in the case of indexicals and demonstratives, especially of pure indexicals like *I, here,* and *now,* it is on the semantic side of the ledger that content varies with context. Consequently, Bach postulates two types of context: narrow and broad context. By *narrow context* is meant any contextual information that is relevant to the determination of the semantic content of, or the assignment of semantic values to, variables such as those concerning who speaks to whom, when, and where. Thus defined, narrow context is semantic in nature. By contrast, *broad context* is taken to be any contextual information that is relevant to the working out of what a speaker overtly intends to covey. It is also relevant to the successful and felicitous performance of speech acts. Hence, it is pragmatic in nature. Given that context plays a role in semantics and pragmatics as well, and that narrow and broad context cannot always be distinguished, the semantics/pragmatics distinction cannot correspond to the context independence/dependence distinction either. In summary, the semantics/pragmatics distinction does not *systematically* coincide with any of the distinctions between truth-conditional versus non–truth-conditional meaning, conventional versus nonconventional meaning, and context independence versus context dependence.

Pragmatic Intrusion Into What Is Said and the Semantics–Pragmatics Interface

On a classical Gricean account of meaning and communication, a distinction is made between what is said and what is (conversationally) implicated. What

is said is generally taken to be (a) the conventional meaning of a sentence uttered with the exclusion of any conventional implicature and (b) the truth-conditional content of the sentence uttered. However, according to Paul Grice, before one computes what is said, one has to (a) identify reference, (b) fix indexicality, and (c) resolve ambiguity. What is implicated is then defined in contrast to, and calculated on the basis of, what is said. It turns out, however, that the determination of (a)–(c) involves pragmatic enrichment of some kind. Putting it another way, it is now generally acknowledged that in working out (a)–(c), there is pragmatic intrusion of some sort into the semantic content of what is said.

Regarding what is said, what the relevance theorists Recanati, Bach, and Stephen Levinson share with Grice is that there is a level of semantic representation or linguistic meaning of a sentence and this level belongs to semantics. Next, what the relevance theorists Recanati, Bach, and Levinson have in common is the view that at least part of the original Gricean notion of what is said has to be understood as involving much more of a pragmatic contribution than Grice has acknowledged. But they cannot agree on two points. The first is that while the relevance theorists, Recanati and Levinson, believe that there is substantial pragmatic intrusion into what is said, as a semantic minimalist, Bach denies that there is such an intrusion. Instead, he posits a level intermediate between what is said and what is implicated, and he calls it the *middle ground of speaker–meaning implicature*. Second, the disagreement concerns the nature of the pragmatic enrichment under consideration. For the relevance theorists, it is an explicature; for Recanati, it is part of the pragmatically enriched said; and for Bach, it is an implicature. In other words, for these scholars, the pragmatic enrichment is of a special kind and differs from conversational implicature. By contrast, on Levinson's view, it is nothing but the same beast as conversational implicature. Consequently, these four analyses of pragmatic intrusion into the classical Gricean characterization of what is said have different implications for delimiting the respective territories of semantics and pragmatics and for the interaction between semantics and pragmatics.

To conclude, semantics and pragmatics constitute two distinct domains of inquiry, but they are inextricably intertwined in such a manner that their borderline is difficult to be drawn in a neat and systematic way.

Yan Huang

See also Language, Philosophy of; Language and Society; Sociolinguistics; Speech Acts

Further Readings

Bach, K. (2004). Pragmatics and the philosophy of language. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 483–487). Oxford, England: Blackwell.

Borg, E. (2007). Minimalism versus contextualism in semantics. In G. Preyer & G. Peter (Eds.), *Context-sensitivity and semantic minimalism: New essays on semantics and pragmatics* (pp. 546–571). Oxford, England: Oxford University Press.

Cappelen, H., & Lepore, E. (2005). *Insensitive semantics: A defence of semantic minimalism and speech act pluralism.* Oxford, England: Blackwell.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition and logical form.* San Diego, CA: Academic Press.

Grice, H. P. (1989). *Studies in the way of words.* Cambridge, MA: Harvard University Press.

Huang, Y. (2007). *Pragmatics.* Oxford, England: Oxford University Press.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature.* Cambridge: MIT Press.

Recaniti, F. (2010). *Truth-conditional pragmatics.* Oxford, England: Oxford University Press.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Oxford, England: Blackwell.

Stanley, J. (2000). Context and logical form. *Linguistics and Philosophy, 23,* 391–434.

Sen's Paretian Liberal

In 1970, Amartya K. Sen christened a simple logical impossibility "the liberal paradox." Ever since, the problem has been discussed in political economy and philosophy. Sen's short paper spread confusion by conceptualizing liberal rights as entitlements to determine *social outcomes* rather than a property of social rules. In short, political (or institutional) liberalism, concerned with liberal forms of the game of politics (game forms), does not suffer from Sen's paradox,

while philosophical (or ethical) liberalism, being concerned with the formation of ethical preferences over social states (liberal welfare functions), does.

Once the rules or "constitution" of a political interaction are defined, outcomes emerge from individual choices on the basis of given preferences. The rules of games—with which political liberalism is concerned—will always allow for an outcome however the rules are fixed (in Table 1, which illustrates a simple game, individuals *can* always choose a row or a column).

Philosophical liberalism scrutinizes the implications of forming ethical preferences over social results in a liberal way. It is not about how to fix the rules of the game but rather about *how* to evaluate the outcomes of games (i.e., how we rank order the cells of Table 1).

A philosophical liberal tries to incorporate the preferences of individuals as far as possible into her own ethical preferences (i.e., her ordering of the cells in Table 1). If an individual A prefers *x* over *y* (cell *x* over cell *y* in Table 1), then the philosophical liberal will intend to "respect" A's ranking when forming her own ethical (or welfare economic) ranking of social states. The obvious problem is that the philosophical liberal cannot at the same time impartially ratify in her ethical preferences B's preference of *y* over *x*.

How this simple point plays out may be illustrated by adapting Sen's original example of a group of two individuals, "Lascivious, A" and "Prude, B," who have diverging preferences over states of the world characterized by their (non)reading of the book *Lady Chatterley's Lover,* or LCL for short. The relevant actions are C_A: "A does *not* read LCL," D_A: "A reads LCL," C_B: "B reads LCL," and D_B: "B does *not* read LCL."

C indicates that the actor is cooperating *with* the *other* one by complying with his other-regarding preferences. D indicates that the actor defects from cooperation with the other-regarding preferences of the *other* actor. The preferences R over social states (the action combinations bringing about the states of the world) are

$$R_A: (D_A, C_B) >_A (C_A, C_B) >_A (D_A, D_B) >_A (C_A, D_B),$$

$$R_B: (C_A, D_B) >_B (C_A, C_B) >_B (D_A, D_B) >_B (D_A, C_B).$$

Prude B prefers that he himself, rather than the lascivious individual, reads LCL. For him, it would be best if neither of the two would have to read the book and worst if both read it. Though the lascivious individual A regards it best if both read LCL, he would volunteer not to read it himself if this should be necessary to induce Prude B to read it. That state of the world is better from the lascivious individual's point of view than one in which only the lascivious person reads the book, and this in turn is better than a state in which no one reads LCL.

If the two actors independently make their choices, then one of the cells of the table will emerge as a result of their actions. Since each of the actors according to his preferences prefers the D alternative, no matter what the other chooses, the (D_A, D_B) result will emerge under independent rational play (the preferences involved are the same as those in the Prisoners' Dilemma). If we identify the action alternatives open to the actors as their respective right of choice, the two actors may want to coordinate how they use their rights to choose between columns or rows independently of each other since $(C_A, C_B) >_A (D_A, D_B)$ and $(C_A, C_B) >_B (D_A, D_B)$. Yet there is no paradox but only a simple collective goods problem. Political liberalism as a mode of assigning rights to act is consistent.

The paradox of (philosophical or ethical) liberalism emerges in the example at hand when a liberal-ethical planner or welfare economist intends to respect the individual preferences of both A and B in her ethical preferences no matter what. If a mapping that ranks social outcome as a function of individual preferences over social outcomes is used to represent the value judgments of such a socio-ethical evaluator, an impossibility result can be proved.

Sen showed that such a mapping cannot conceivably meet three liberal requirements simultaneously:

1. The preferences of each individual are at least over one pair of social outcomes decisive for the

Table 1 Reading *Lady Chatterley's Lover*

B	C_B	D_B
A	*B reads*	*B reads not*
C_A, A reads not	(C_A, C_B)	(C_A, D_B)
D_A, A reads	(D_A, C_B)	(D_A, D_B)

Source: Author.

ethical preferences of the planner ("minimal liberalism," or "liberal decision right").

2. An alternative to which another one exists that is preferred by all should not be ethically preferred by the planner ("weak Pareto principle," or "social efficiency").
3. Whatever the preferences of the individuals over the social alternatives are, they will be taken into account as ethically relevant by the preference-forming mechanism ("unrestricted domain," or "individual sovereignty").

Using the preference orders of Table 1, it is obvious that it is *impossible* to ratify all preferences *over the cells* of the table simultaneously in the ethical liberal's welfare judgment and to secure Pareto optimality. Yet the problem emerges obviously quite independently from the Pareto principle.

Hartmut Kliemt

See also Collective Rationality; Decision Theory; Libertarianism, Political; Pareto Optimality; Preference; Rational Expectations; Social Choice Theory; Welfare Economics

Further Readings

Buchanan, J. M. (1996). An ambiguity in Sen's alleged proof of the impossibility of a Pareto liberal. *Analyse & Kritik, 18*(1), 118–125. (Original work published 1975)

Sen, A. K. (1982). *Choice, welfare, and measurement.* Oxford, England: Oxford University Press.

Suzumura, K., & Suga, K. (1986). Gibbardian libertarian claims revisited. *Social Choice and Welfare, 3*(1), 61–74.

SERENDIPITY

In philosophy, the idea of serendipity goes mostly by the title of "accidental discovery," while in the social sciences, *serendipity* is the preferred term. Whichever the nomenclature, the idea has been at most a relatively minor concern in the two fields, explained in part by its roots in a fairy tale (see final paragraph of this entry) and in part by the absence of a long-standing intellectual interest notably pursued by recognized scholars. It is, however, an important, even if unduly neglected, notion in scientific practice. It is also linked to recent attempts to study creativity.

The sociologist Robert Stebbins defined *serendipity* as the quintessential form of informal experimentation, accidental discovery, and spontaneous invention. As a method of discovery, serendipity joins several others, among them discovery by exploration, trial and error, and metaphoric reasoning. Robert K. Merton (who studied serendipity most systematically) and Elinor Barber have observed that serendipity can refer either to finding something of value while searching for something else or to finding something sought after in an unexpected place or manner.

In the social sciences, serendipity as a method of discovery comes closest to exploration and is sometimes confused with it. Exploration is a broad-ranging, purposive, systematic, prearranged undertaking. Moreover, serendipity as its usage in popular culture suggests is conceivably available for everyman, whereas exploration attracts a much more exclusive band of devotees. It is the department of those creative people who are trained to routinely produce new ideas. In certain fields of serious leisure and professional work, artists, scientists, and entertainers, for example, routinely explore, while in some forms of casual leisure, people at play (both children and adults), sociable conversationalists, and seekers of sensory stimulation rarely if ever do so. The latter set usually makes its discoveries only through serendipity. By contrast, though the routinely creative group occasionally discovers something serendipitously, it is far more likely to rely on exploration for new ideas.

Merton introduced the idea of serendipity to the social sciences. An anomalous finding during his research in the 1940s on the social organization of the suburban American community of Craftown stirred his interest in the unexpected in science. Craftown was a predominantly working-class community. There he observed that a sizable proportion of its residents were affiliated with more civic, political, and other volunteer organizations than where they had lived previously. He noted further and serendipitously that this increased group participation had occurred primarily among the parents of infants and young children. This finding ran counter to commonsense knowledge, for it was widely believed that, particularly in the lower socioeconomic levels, youngsters tend to tie parents down. Thus constrained, they are unable to participate in organized group life outside the home.

Although serendipity is not exploration, it can occur through inductive generalization. Researchers in the social sciences collecting data using a controlled study design sometimes observe regularly occurring phenomena that are fully unexpected. This is what happened during Merton's study while he was searching for something else. In such circumstances, the controls were, obviously, incomplete; despite the best intentions, a significant recurrent extraneous phenomenon found its way to Merton's attention. Furthermore, he was alert enough to recognize its importance for his investigation. This, since it is based on recurring instances, is *inductive serendipity.*

What may be called *unique serendipity* is probably the more common of the two types. Everyday life offers many examples: seeing a species of birds in a region they are believed not to inhabit; noting the improved flavor of a soup after accidentally adding a spice not called for by the recipe; and recognizing the greater efficiency of a new (auto, bicycle, pedestrian) route to the office upon being forced to abandon the old one because of repairs. Stebbins argues that one of the benefits of pleasurable reading lies in its serendipitous revelations. In unique serendipity, the discovered bird, flavor, or efficiency is a single occurrence, albeit with subsequent occurrences serving to support the conclusion about the significance of the discovery. That is, other people see the same bird in the same region, the soup continues to taste better enhanced with the new spice, and the new route to the office continues to be more efficient than the old one. *Unique serendipity* is a label for what Merton and Barber indicate is "finding something sought after in an unexpected place or manner."

Inductive serendipity might be described as the same process as exploration were it not that the former is accidental whereas the latter is planned. Both processes, when implemented in research, result in generalizations that emerge from systematic, direct observation of phenomena. But those resulting from exploration are intentionally sought by way of a study design that facilitates discovery. Not so with inductive serendipity.

Both exploratory and serendipitous generalizations emerging as inductively generated conclusions take their significance from a larger explanatory framework of some kind. The same holds for unique serendipity. Social-scientific theory, including other grounded theories from the exploratory project itself, helps exploratory investigators interpret what they have observed. In the case of unique serendipity, people experiencing such discoveries interpret them according to a related fund of knowledge and experience. The bird is identified as new to the region according to the birder's knowledge of regional avian wildlife. The cook, based on past culinary involvements, asserts that the new flavor is superior. And so on. It is against this kind of background that serendipitous discoveries derive their meaning.

Recently, in a far more detailed examination of serendipity than was found in Merton's earlier study, Merton and Elinor Barber explored its nature, history, and application in the humanities and social sciences. They observed that *serendipity* as a term first appeared in a letter written in 1754 by Horace Walpole to Horace Mann. Walpole coined the word *serendipity* based on his familiarity with the fairy tale *The Travels and Adventures of Three Princes of Serendip*. Serendip is the ancient name of Ceylon (today Sri Lanka), and the three princes were sons of Jafer, at the time philosopher-king of that country. Serendipitous discovery is evident in places in the tale.

Robert Alan Stebbins

See also Abduction and Inference to the Best Explanation; Deduction; Explanation, Theories of; Induction and Confirmation

Further Readings

Merton, R. K. (1957). *Social theory and social structure* (Rev. ed.). Glencoe, IL: Free Press.

Merton, R. K., & Barber, E. (2004). *The travels and adventures of serendipity.* Princeton, NJ: Princeton University Press.

Stebbins, R. A. (2001). *Exploratory research in the social sciences*. Thousand Oaks, CA: Sage.

Stebbins, R. A. (2013). *The Committed Reader: Reading for utility, pleasure and fulfillment in the twenty-first century.* Lanham, MD: Scarecrow Press.

SEXUALITY

Sexuality as a word was developed to describe a host of biological phenomena. It has also taken on a more specific meaning, indicating an individual's sexual proclivities or identity: "my sexuality." More recently, however, it has been increasingly used to

refer to the historical organization and cultural significance of erotic behavior. This points to the growing importance of sexuality in political debate on a globalizing world. It has become a crucial aspect of fundamentalist politics, and it has also given rise to a new concept of human sexual rights.

This entry presents the historical, political, scientific, and social-scientific aspects of the notion and underlines links with epistemological issues as well as with philosophical questions regarding biological kinds and essentialism.

Historical Overview

The history of the term suggests its complex evolution. The earliest usage of the word *sex* in the 16th century referred to the division of humanity into the male section and the female section. The subsequent meaning current since the early 18th century refers to physical relations between the sexes, "to have sex." *Sexual,* a word that can be traced back to the mid 17th century, carries similar connotations—pertaining to sex, or the attributes of being male or female, is one given meaning. The term *sexuality* itself emerged early in the 19th century, meaning the quality of being sexual. It is this meaning that is carried forward and developed by sexologists, who emerged as a distinct category of specialists in the late 19th century and became increasingly influential in the 20th century—luminaries such as the Austrian pioneer Richard von Krafft-Ebing, the British writer Havelock Ellis, the German author Magnus Hirschfeld, and the most famous of them all, Sigmund Freud, the founder of psychoanalysis.

Epistemological and Metaphysical Issues

Sexologists sought to discover the "laws of nature," the true meaning of sexuality, by exploring its various guises and manifestations. They all concurred that sexuality was in some ways an underlying biological essence that underlay a range of human behavior and fantasies. They sought to classify and define the various forms they discovered, and it is to this generation that we owe concepts and terms such as *homosexuality* and *heterosexuality, transvestism* and *sadomasochism, coprophilia* and *necrophilia,* and a thousand more terms in the ever-growing encyclopedia of sexual naming. Later researchers such as Alfred Kinsey and his colleagues demonstrated that sexual behavior was much more diverse than public morality and private prejudice liked to believe, and they argued that in biology, there was neither right nor wrong and neither normal nor abnormal.

To this point, sexuality was seen largely as a biological phenomenon, with crucial psychic manifestations. Since the 1960s, however, contemporary theorists have questioned the naturalness and inevitability of the sexual categories and assumptions people have taken for granted. They suggest that the sexologists did not so much discover or map the world of sexuality as help create and constitute it. The concept of sexuality, they argue, unifies a host of activities that have no necessary or intrinsic connection: discourses, institutions, laws, regulations, administrative arrangements, scientific theories, medical practices, household organization, subcultural patterns, ethical and moral practices, and the arrangements of everyday life. Nothing is sexual, the British sociologist Ken Plummer suggested, but naming makes it so. So sexuality was now seen as a narrative, a complexity of the different stories we tell each other about the body; a series of scripts through which we enact erotic life; or an intricate set of performances through which the sexual is invented, ritualized, and embodied. It is the name, Michel Foucault famously argued, of a historical construct.

Historical, Political, and Cultural Approaches

So instead of seeking the laws of nature that would explain sexuality as a universal phenomenon, the new theories suggested that we need to examine the specific historical and cultural organizations of sexualities in all their diversity. We should speak not of sexuality but of sexualit*ies,* not of sex and society but of sexual cultures. A further point follows from this: a recognition that sexualities are hierarchically organized, with some forms dominant while others have been historically subordinate and marginalized, shaped by complex relations of power. The most familiar of these relate to gender, with male domination shaping attitudes to female sexuality and homosexuality. Hierarchies around age, class, race, and ethnicity have also been crucial in shaping sexual beliefs and behavior. In recent years, there has also been an increasing recognition that sexualities, at least in the West, have been organized into institutionalized forms of heterosexuality.

Power relations structure sexuality, but uneven power relations also provide the necessary condition

for resistance and for sexual politics. The impact of feminism and of lesbian, gay, bisexual, and transgender (LGBT) politics since the late 1960s has simultaneously sharpened awareness of the arbitrary nature of gender and sexual categories and increased the possibility of changing them. New sexual identities, like LGBT, have become the basis for radical claims to equality.

At the same time, there has been a growing recognition that the meanings developed around sexual identity in the West were not directly relevant to many marginalized people both in Western countries and across the globe. This is closely related to a recognition of the growing importance of globalization in relation to the organization of sexualities. The spread of the HIV/AIDS epidemic since the 1980s to become a worldwide pandemic is a tragic illustration of this. Sexuality has become a major element in global political discourse, with new forms of conflict—over, for example, the rights and roles of women, attitudes toward homosexuality and gender nonconformity (e.g., transgender)—integral to the emergence of fundamentalist politics. At the same time, a new language of human sexual rights has emerged that provides a framework through which a necessary dialogue can develop on how to recognize and live with each other's differences and a common humanity.

Jeffrey Weeks

See also Essentialism; Feminism: Schools of Thought; Foucault's Thought; Identity, Social; Kinds: Natural Kinds Versus Human Kinds; Love, in Social Theory; Love, Philosophy of; Social Constructivism; Symbolic Interactionism

Further Readings

Altman, D. (2001). *Global sex*. Chicago, IL: University of Chicago Press.

Ellis, H. (1911). *Studies in the psychology of sex* (Vols. 1–6). Philadelphia, PA: F. A. Davis.

Foucault, M. (1979). *The history of sexuality: Vol. 1. An introduction* (R. Hurley, Trans.). London, England: Allen Lane.

Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. (1948). *Sexual behavior in the human male*. Philadelphia, PA: W. B. Saunders.

Plummer, K. (1979). *Telling sexual stories: Power, change and social worlds*. New York, NY: Routledge.

von Krafft-Ebing, R. (2011). *Psychopathia sexualis* (S. K. Franklin, Trans.; Complete Authorized English Translation). New York, NY: Arcade. (Original work published 1886)

Weeks, J. (2009). *Sexuality* (3rd ed.). New York, NY: Routledge.

Simmel's Philosophy of Society

Georg Simmel (1858–1918) is a thinker in whom the dialectic of the relationship between philosophy and the social sciences is reflected perhaps in its most refined and complex form. Nowadays, Simmel is generally perceived as one of the classics of sociology, alongside Max Weber, Ferdinand Tönnies, and Émile Durkheim. Yet Simmel did not consider himself a sociologist and even claimed on many occasions that sociology was for him merely a secondary activity. Although he published a great number of works in this field and was among the founders of the German Society for Sociology, he regarded himself (and was regarded by contemporaries) mainly as a philosopher.

This determined the character of Simmel's contribution to the social sciences. It was the place of the social and of our knowledge of the social on the map of human activity, and the human spirit in general, that preoccupied him most. Simmel was less interested in thorough and systematic empirical sociological research. His writings are full of specific observations and examples, which relate to even the minutest details of human life and interaction, yet they generally serve as illustrations of some general line of thought and are derived either from our common everyday experience or from the research findings of other scholars.

Simmel's Early Sociology

Simmel's early sociological writings, among them his first book, *On Social Differentiation* (1890), are written in the tradition of early anthropology, *Völkerpsychologie* (people's psychology), and Spencerian evolutionary sociology. The underlying motif of these works is the process of social evolution, which leads to the formation of modern, highly developed civilization. Simmel argues that modern society is characterized by growing differentiation,

leading to the development of a modern individuality, on the one hand, and to the strengthening and widening of social networks, interactions, and circles, on the other. This view is akin in many respects to those of many other classical sociologists, such as Herbert Spencer, Tönnies, and Durkheim. Simmel also shares with these thinkers, especially with Durkheim, a certain sociological reductionism. All aspects of human life and spirit appear in it as reducible to and explainable in terms of their meaning for social life. At the same time, even in his early writings, this sociological reductionism has its limits. Simmel never fully subjects spheres such as religion, art, and philosophy exclusively to sociological explanation. And in his later writings, he discards sociological reductionism even more decisively.

Formal Sociology

This turn is signified by two seemingly opposite attitudes toward the science of society that Simmel developed in his mature period. On the one hand, he completed his major work in the field—*Sociology* (1908)—supposedly his definitive statement of what sociology should be. On the other hand, this is paralleled by his reduced interest in all things social, as he turned his attention to other intellectual fields, stating occasionally that he was a philosopher rather than a sociologist. But there is no contradiction here: It is precisely the renunciation of the imperialistic claims of sociology that allowed Simmel to delineate more narrowly and, consequently, more precisely its proper field.

Simmel's vision of the proper sphere of sociology is first outlined in the 1894 essay "The Problem of Sociology." According to him, sociology must be the science of social interaction as such. Many other disciplines (e.g., economics, ethics, and jurisprudence) are indeed also related to the study of social interactions. Yet they investigate these interactions in terms of their content, which may include meanings, consequences, and so forth. Sociology, by contrast, must study the social *forms* implied in interaction. The same form may be present in social interactions of different kinds, and conversely, the same interaction may entail the whole variety of forms. For example, the form of the quantitative relation between three actors may produce similar dynamics in spheres as different as family, economic life, or international relations. Conversely, the dynamics of relationships in a certain family may exemplify a variety of forms of interaction, with respect to the number of persons involved, the structure of the inner hierarchy, the spatial relations, and so forth.

Formalism was already a familiar approach in other disciplines, such as "formal" psychology or "formal" aesthetics. Simmel appears to have been the first to introduce it in the field of sociology. His *Sociology* was an attempt to show what formal sociology may look like. Yet the book, despite its length, did not offer a complete scheme of those forms, remaining rather a collection of exemplifications of the suggested approach. Simmel himself regarded this work not as a final statement but as an illustration of fruitful directions of research in the sociology of forms.

The Legacy of Simmel's Sociology

The story of Simmel and his sociology is full of paradoxes. He was the most philosophically minded classical social thinker, yet he dethroned sociology of its philosophical claims and assigned to it a strictly scientific arena. He envisioned the future science of society in most rigorous terms, yet his own sociological works fall short of this rigor and betray him as a philosopher and essayist rather than a systematic scientist. He enjoys respect in today's sociology, usually not because of his general formal method but rather on account of his numerous essay-like inquiries into multiple aspects of modern life. Simmel is thus valued as a sociologist but not as the founder of formal sociology. Two different explanations can account for this. Some would (and often did) claim that without meaning and intentionality, the study of society would lose any significance and turn into a meaningless play of formal structures. But it is also possible that Simmel offered us a glimpse into an ideal science of society that still cannot be practiced with the tools available within the discipline. If that is the case, formal sociology may still be waiting for its great practitioner.

Efraim Podoksik

See also Durkheim's Philosophy of Social Science; Modernity; Neo-Kantianism; Philosophy of Sociology, History of; Scheler's Social Person; Weber and Social Science: Methodological Precepts

Further Readings

Frisby, D. (2002). *Georg Simmel.* London, England: Routledge.

Gerhardt, U. (2006). The two faces of modernity in Georg Simmel's social thought. *Simmel Studies, 16*(1), 49–74.
Levine, D. N. (1959). The structure of Simmel's social thought. In K. H. Wolff (Ed.), *Essays on sociology, philosophy and aesthetics* (pp. 9–32). New York, NY: Harper & Row.
Simmel, G. (1972). *On individuality and social forms* (with an introduction by D. Levine, Ed.). Chicago, IL: University of Chicago Press.
Simmel, G. (2009). *Sociology: Inquiries into the construction of social forms* (Vol. 1). Leiden, Netherlands: Brill. (Original work published 1908)
Spykman, N. J. (1965). *The social theory of Georg Simmel.* New York, NY: Atherton Press.
Weingartner, R. H. (1962). *Experience and culture: The philosophy of Georg Simmel.* Middletown, CT: Wesleyan University Press.

Simulation Theory

This entry introduces one of the alternative theories of mind, Simulation Theory; contrasts it with rival views; and presents recent developments in child psychology and neuroscience. The entry goes on to point out the resemblance of modern Simulation Theory to, and its difference from, earlier hermeneutic accounts of social sciences or classical theories of empathy.

Simulation Theory was originally developed as an alternative to a long-dominant view in philosophy of mind that our everyday understanding of human behavior is based on a tacit body of general knowledge, roughly comparable to a scientific theory, of the workings of the human mind. The alleged theory—often called *folk psychology*—posits unobservable mental states such as beliefs, desires, intentions, and feelings, linked to each other and to observable behaviors by lawlike principles. These principles, supposedly shared by people of all cultures and virtually all levels of intelligence, are applied to observable situations by way of logical inferences that generate predictions and explanations of behavior.

According to Simulation Theory, rather than having to inherit or acquire a body of general theoretical information about the mind (vide "Theory Theory"), human beings are able to use the resources of their own minds to simulate the mental states and processes of others by *mental simulation*—that is, to generate similar states and processes in themselves.

A typical method, though not the only one, would be role-taking, or "putting oneself in the other's place," and making decisions within the context of such pretense. For example, one might anticipate the product of another's theoretical or practical inferences from given premises by making inferences from the same premises oneself; or knowing what the product is, one might retroduce the premises, perhaps by something resembling unconscious Bayesian inference. To reason from the same premises would typically require indexical adjustments, such as shifts in spatial, temporal, and personal "points of view," to place oneself in the other's physical and epistemic situation insofar as it differs from one's own. One may also compensate for the other's reasoning capacity and level of expertise, if possible, or modify one's character and outlook as an actor might, to fit the other's background and behavioral history. Such *adjustments,* even when insufficient for making decisions in the role of the other, allow one to discriminate between action options likely to be attractive to the agent and those likely to be unattractive. Accordingly, one would be prepared for the former actions and surprised by the latter.

Simulation Theory and Theory Theory need not be formulated in ways that exclude one another. The resources underlying our everyday psychological understanding, or what psychologists (and others) call *theory of mind* (ToM), might incorporate both simulational and theoretical components. Indeed, a number of philosophers and psychologists have put forward *hybrid accounts,* usually with the suggestion that people rely on simulation in some contexts and theory in others.

Developmental Psychology

The theory versus simulation debate within philosophy of mind has had an important impact on research by developmental psychologists on the *child's* theory of mind. The attribution of false belief, in particular, might be conceived in two ways: (1) the young child comes to posit an unobservable internal state, belief, as intervening between the other's external situation and his behavior or (2) the young child comes to develop a new skill, that of introducing into his simulation of the other a counterfactual pretend world. Although the issue has motivated a considerable amount of empirical research, it remains unresolved.

Mirror Neurons

The discovery in the early 1990s of mirror neurons in the primate brain opened another channel of interdisciplinary research and debate. Certain neuron systems that are involved in the production of our own actions and emotions also become active when we observe similar actions or emotions in another person. A major research topic in the new discipline of social neuroscience concerns the possible contribution of mirror systems to our psychological understanding of others. Whether and how these "low-level" systems might facilitate the kind of "high-level" mental simulation that philosophers and psychologists have been discussing remain open questions.

Earlier Views: Similarities and Differences

In some respects, Simulation Theory resembles the views of earlier writers, such as the German philosopher Wilhelm Dilthey and the British idealist R. G. Collingwood, on the role of empathetic understanding (*Verstehen*) and historical reenactment in the human sciences, as well as suggestions by Immanuel Kant and Willard van Orman Quine. However, unlike these earlier theorists, simulation theorists often appeal to empirical findings, particularly experimental results in developmental psychology and neuroscience.

They also encourage speculation grounded in neuroscience about the computational mechanism that might accomplish the task of simulation: presumably one that calls up resources ordinarily used for engagement with the world but runs them "offline," so that their output is not actual behavior but only predictions or anticipations of behavior and their inputs and system parameters are accordingly not limited to those that would regulate one's own behavior.

Robert M. Gordon

See also Developmental Psychology; Empathy; Folk Psychology; Mirror Neurons and Motor Cognition in Action Explanation; *Naturwissenschaften* Versus *Geisteswissenschaften;* Theory Theory

Further Readings

Carruthers, P., & Smith, P. K. (Eds.). (1996). *Theories of theories of mind.* Cambridge, England: Cambridge University Press.

Dokic, J., & Proust, J. (Eds.). (2002). *Simulation and knowledge of others.* Amsterdam, Netherlands: John Benjamins.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading.* Oxford, England: Oxford University Press.

Gordon, R. M. (1986). Folk psychology as simulation. *Mind and Language, 1,* 158–171.

Gordon, R. M. (2000). Simulation and the explanation of action. In K. R. Stueber & H. H. Kogaler (Eds.), *Empathy and agency: The problem of understanding in the human sciences* (pp. 62–82). Boulder, CO: Westview Press.

Situated Action

Situated action (SA) is one of a set of theories in cognitive science that puts primacy on the role of context in understanding cognition. Like related theories labeled *situated cognition* and *situated learning,* foundational across these is the notion that the environment is an essential element *of* cognition.

SA is most associated with the social anthropologist Lucy Suchman, who proposed this theory to argue that cognition cannot be separated from the agent and the environment, whether it be the physical, social, or cultural environment. SA owes its scholarly roots to the anthropological and sociological accounts of culture and cognition, for example, of Max Weber, as well as the educational philosophy of John Dewey. SA posits an integrated or holistic view of cognition, where one is not *in* an environment but *part of* an environment. Against this broader theoretical backdrop, SA was developed to empirically address what Suchman referred to as "the irreducibility of lived practice, embodied and enacted." What differentiated SA was a focus on plans and their relation to cognition as it emerges in context. In SA theory, plans often bear little resemblance to the actions that follow and were more likely post hoc rationalizations for action. SA drew from ethnomethodology and the study of human activity via observations, as it arises *between* a person and the setting in which that activity takes place.

Studying Cognition in Context

SA was made most prominent by Suchman's 1987 book *Plans and Situated Action* (P&SA), written to codify her experiences studying human–machine interaction while at Xerox PARC. P&SA was meant as an alternative to prevailing views within artificial intelligence (AI) that knowledge could be adequately

represented within a system such that it could predict and manage interactions that had yet to be experienced. In particular, P&SA was a reaction to AI's overreliance on knowledge representations and control structures devised to execute goals and plans in service of intelligent interactive interfaces. SA theory proposed that such approaches could not account for the relational coupling that emerges during interaction between a human and a machine. Rather, what is required is a view that overcomes asymmetries between humans and machines as interactional partners, such that machines are able to more accurately perceive human actions as they actually occur.

Foundational to understanding SA is the role of plans in human activity. Within SA accounts, plans are not cognitive control structures that determine action. Because activity can only emerge from the interaction of the human and a situation, plans cannot be considered as fully structuring activity. Rather, planning is a type of imaginative practice producing projections that may relate to future actions. More broadly, SA was a reaction to the rational theories of problem solving prevalent in AI and cognitive science, which focused on knowledge residing within the mind of a problem solver—knowledge that was seen as deterministic of behavior. SA theory argued that plans do not fully shape behavior and that activity was more adaptive and flexible. Furthermore, SA pointed to the important role that the environment played in determining action. Thus, SA did away with the dualistic view of mind and body to draw attention to the interaction of mind, body, and environment, where plans serve as resources for action but do not fully determine activity.

Critiques of SA

SA and associated theories broadly influenced cognitive science and how cognition should be understood and studied. This influence was felt in disciplines ranging from AI to the learning sciences. SA theory was influential enough to spawn special issues in two journals, *Cognitive Science* (1993) and the *Journal of the Learning Sciences* (2003). Along with this influence came criticism from a number of perspectives. From traditional views of cognitive science and theories based on symbolic accounts of cognition, Alonso Vera and Herbert Simon argued from the information-processing perspective. They noted that symbolic theories of cognition have both substantial empirical evidence as well as the means to account for context and situated activity. An additional criticism centered on the interpretation that, as a theoretical approach, SA rejected the utility of planning entirely. From the learning sciences, researchers took issue with some of the foundational claims of situated theories concerning context and specificity of task. John Anderson and his colleagues drew from research in cognitive psychology to counter claims from SA theory and argued that abstract instruction has been found effective, that the teaching of skills in the absence of context has been documented, and that transfer of learning across contexts has been empirically verified.

From computer science and understanding human–computer interaction (HCI), others argued that SA was unable to account for the broad range of human activity and, thus, had limited usefulness as a theory. By focusing too narrowly on the interactions that emerge in specific situations, comparisons across contexts were too challenging. Rather, what was required was a means of abstracting key features of interaction that occur over many domains, and the artifacts used within domains, such that they can inform HCI design. Blending these critiques, Bonnie Nardi argued that "activity theory," emerging out of work analysis in the former Soviet Union, provides a superior means of addressing the role of context in cognition. Building on the work of Alexei Leontiev, Nardi argued that SA is unable to account for the durable and stable phenomena that consistently emerge across situations. Activity theory, though, shifts the focus from the context to the activity itself and to how artifacts mediate activity. In this way, even when conditions change, the activity is focal, and what matters is how goals and operations are modified to adapt to the conditions but still enable completion of the activity. By acknowledging that goals shape activity, as opposed to simply being post hoc descriptions of activity, Nardi argued that one can more fully account for the design of artifacts in support of goal-driven interaction.

Lasting Impact of SA

In later writings, SA theorists argued that what SA theory was meant to accomplish was to move the field toward a true understanding of the relationship between plans and the actual activity that ensued. In the 2007 follow-up edition to P&SA, Suchman noted that her initial objective was for research to move away from the view of plans as deterministic to one where plans became the actual objects of

study. Many were sympathetic to this conceptualization of SA and credit these ideas with influencing cognitive science and related fields such as HCI design and the learning sciences. William J. Clancey was a strong advocate for situated theories in cognitive science and argued against notions of stored symbols or representations in memory to suggest that situated actions are more akin to emergent neurological coordination. The overall influence of SA led the learning scientist James Greeno to propose the term *situativity* in order to acknowledge that all cognition and action are situated and that research needs to understand this relational coupling where meaning is constructed within particular contexts. The lasting impact of SA theory has been the recognition of the need to analyze the execution of actions in context so as to understand the often spontaneous adaptation of response to differing or unforeseen aspects of the situation. This, it was argued, should be the focus of how we come to understand human cognition, in general, and human–technology interactions, in particular.

Stephen M. Fiore

See also Artificial Intelligence; Cognitive Sciences; Ethnomethodology; Goal-Directedness; Human–Machine Interaction; Situated Cognition

Further Readings

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5–11.

Chaiklin, S., & Lave, J. (Eds.). (1993). *Understanding practice: Perspectives on activity and context*. Cambridge, England: Cambridge University Press.

Clancey, W. J. (1997). *Situated cognition: On human knowledge and computer representations*. Cambridge, England: Cambridge University Press.

Dewey, J. (1938). *Experience and education*. New York, NY: Simon & Schuster.

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist, 53*(1), 5–26.

Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life*. Cambridge, England: Cambridge University Press.

Nardi, B. (Ed.). (1996). *Context and consciousness: Activity theory and human-computer interaction*. Cambridge: MIT Press.

Norman, D. A. (Guest Ed.). (1993). Situated action [Special issue]. *Cognitive Science, 17*(1).

Robbins, P., & Aydede, M. (2008). *The Cambridge handbook of situated cognition*. Cambridge, England: Cambridge University Press.

Suchman, L. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge, England: Cambridge University Press.

Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge, England: Cambridge University Press.

*The following is a collection of commentaries on Suchman's 1987 book (*Plans and Situated Actions*) and her response to it:*

Koschmann, T. (2003). Plans and situated actions: A retro-review. *Journal of the Learning Sciences, 12*(2), 257–258. doi:10.1207/S15327809JLS1202_4

Sharrock, W., & Button, G. (2003). Plans and situated action ten years on. *Journal of the Learning Sciences, 12*(2), 259–264. doi:10.1207/S15327809JLS1202_5

Wells, G. (2003). Lesson plans and situated learning-and-teaching. *Journal of the Learning Sciences, 12*(2), 265–272. doi:10.1207/S15327809JLS1202_6

Carroll, J. M. (2003). Situated action in the zeitgeist of human-computer interaction. *Journal of the Learning Sciences, 12*(2), 273–278. doi:10.1207/S15327809JLS1202_7

Vera, A. H. (2003). By the seat of our pants: The evolution of research on cognition and action. *Journal of the Learning Sciences, 12*(2), 279–284. doi:10.1207/S15327809JLS1202_8

Schwartz, D. L., & Martin, T. (2003). Representations that depend on the environment: Interpretative, predictive, and praxis perspectives on learning. *Journal of the Learning Sciences, 12*(2), 285–297. doi:10.1207/S15327809JLS1202_9

Suchman, L. (2003). Writing and reading: A response to comments on plans and situated actions. *Journal of the Learning Sciences, 12*(2), 299–306. doi:10.1207/S15327809JLS1202_10

Situated Cognition

The situated perspective on cognition encompasses a whole host of ideas relevant to the study of mind in the social sciences. Among the more familiar of these ideas is the principle of *environmental embedding*. According to this principle, cognitive activity depends so heavily upon an agent's interaction with the natural and social environment that it cannot be understood in abstraction from that context.

Cognition is essentially an ecological phenomenon and must be studied as such.

A second, more controversial, aspect of the situated perspective is the principle of *environmental extension.* This is the idea that the boundaries of cognition need not coincide with the bodily boundaries of the organism, since an agent's cognitive processes may literally extend into the environment—not just in principle but in practice. Though these two ideas about the context dependence of cognition—embedding and extension—are but two of many variations on a broad theme, their centrality to the situated paradigm makes them good focal points for an overview of the topic. This entry briefly explores each of them in turn.

The Embedded Mind

Perhaps the best-known exemplar of the ecological approach to cognition is J. J. Gibson's landmark work on vision, *The Ecological Approach to Visual Cognition.* On Gibson's view, vision requires the movement of an organism through the environment, as it is this movement that enables the organism to extract important information contained in the *ambient optic array,* the structure of light that impinges on the perceiver. For Gibson, the starting of vision is not the passive processing of the retinal image, something that is internal to the organism, but rather the active processing of the optic array, a part of the environment. This processing also underwrites the efficient detection of *affordances,* agent-relative functional properties of objects in the lived environment (e.g., the "climb-ability" of a ladder, the "sit-on-ability" of a chair).

Another facet of the ecological perspective on cognition involves the phenomenon of informational off-loading from organism to environment. As David Kirsh has observed, people routinely perform a variety of *epistemic actions,* that is, they modify the environment so as to facilitate problem solving by simplifying the computational task. He gives the example of a grocery bagger, who sorts items by their task-relevant characteristics (e.g., weight, bulk, and fragility) into different partitions of a spatially articulated buffer zone before packing them into bags. By structuring the local environment in this fashion, the bagger lowers her internal informational overhead—especially the load on working memory, with its severely limited capacity—simplifying the decision-making process and boosting task performance. Other examples of epistemic action come from the domain of game play. Most Scrabble players, for instance, explore the space of possible words by physically rearranging the letter tiles on the rack in front of them rather than in their heads. The general principle is this: Cognitive agents do best when they minimize internal representation and processing by using "the world as its own model" (in Rodney Brooks's nice phrase), and epistemic actions are a good way of accomplishing that minimization.

Attention to the embeddedness of mind is likewise evident in ecologically oriented research on human rationality, in general, and on judgment and decision making, in particular. The "fast and frugal heuristics" program, for example, draws heavily on the idea that the utility of simple decision-making rules depends on their fit with the environmental context. Thus, a rule that uses recognition as a guide to some criterion of interest (e.g., the relative population size of cities) will yield good results only in settings where recognition is strongly positively correlated with that criterion. A similar concern with ecological validity can be found in decision-field theory, which emphasizes the dynamic, temporally constrained, and interactive character of decision making, as well as the influence of context on preference orderings. Finally, much of social psychology, notably research in the "situationist" tradition of Lee Ross and Richard Nisbett's book *The Person and the Situation,* is concerned with the ways in which various aspects of the social environment (e.g., the communicative context, the context of personal relationships, and the context of group membership) shape our thought and action.

The Extended Mind

A more radical variation on the situated cognition theme is Andy Clark and David Chalmers's idea that cognitive states and processes literally extend into the environment. This revisionary metaphysics of mind is motivated, at least in part, by epistemological considerations of the sort mooted earlier: namely, that the deep embeddedness of mind in the world dictates that we cannot understand how the mind works without taking into account how mind and world interact. The view of the mind as a complex dynamical system, a view descended from cybernetics and control theory, also pulls in

this direction, insofar as it emphasizes causal interactions between the organism and its environment and de-emphasizes the boundary between them. But arguably, the strongest support for the extended-mind hypothesis comes from the results of clever thought experiments. The celebrated case of Otto, an Alzheimer's patient who regularly deploys information stored in a notebook rather than information stored in biological memory, is instructive here. From a functionalist perspective on cognition at least, argue Clark and Chalmers, Otto's externally encoded "memories" are the genuine item, and the differences between his cognitive profile and that of a neurotypical agent, with biological memory intact, are superficially significant at best. As for the fact that Clark and Chalmers's proposed redrawing of the boundary between mind and world runs afoul of commonsense intuitions, they say, "So much the worse for common sense."

Criticism

Whether this radicalism can be sustained is an open question. Certainly, it has its detractors. Most notably, it has been argued that the conceptual shift from embeddedness to extension rests on conflating the distinction between causation and constitution. It has also been argued that revising the metaphysics of mind in the manner suggested would have dire consequences for cognitive science, for example, rendering most if not all nomological generalizations about memory (e.g., recency and primacy effects, interference effects, etc.) null and void. But the jury is still out on this subject—and, given the history of metaphysical debates in general, is likely to remain so for some time to come.

Philip Robbins

See also Cognitive Sciences; Distributed Cognition and Extended-Mind Theory; Embodied Cognition; Evolutionary Psychology; Grounded Cognition and Social Interaction; Group Mind; Situated Action; Social Perception

Further Readings

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence, 47,* 139–159.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58,* 10–23.

Gibson, J. J. (1986). *The ecological approach to visual perception.* New York, NY: Psychology Press.

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence, 7,* 31–68.

Robbins, P., & Aydede, M. (Eds.). (2009). *The Cambridge handbook of situated cognition.* Cambridge, England: Cambridge University Press.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation.* New York, NY: McGraw-Hill.

Situational Analysis

Situational analysis (SA) is a recent extension of Straussian grounded theory (GT) for qualitative data analysis in social sciences and related research. It is also inspired by Donna Haraway's concept of situated knowledges, Foucauldian discourse analyses, and social studies of science and technology. The focus is on analysis of the situation being researched: the elements in it, the relations among elements, the conditions of possibility for action, and related discourses. SA integrates poststructural assumptions, feminist emphases on differences and complexities, and analysis of power and encourages studying narrative, documentary, historical, and visual discourses. The situation of inquiry itself, broadly conceived, is the unit of analysis.

Philosophical Underpinnings of Situational Analysis

Like GT, SA is also grounded in symbolic interactionism, a sociological tradition with roots in American pragmatist philosophy (George Herbert Mead, John Dewey) and Charles S. Peirce's concept of abduction as generative of theorizing. The fundamentals of symbolic interactionist theory, constructionist GT, and SA constitute a "theory–methods package." Ontology and epistemology are ultimately nonfungible. What can be known and how we can know it are inseparable. Such packages also include the concrete practices through which social scientists go about their work.

What is new in SA? There are three key facets. First, early-20th-century Chicago social ecologies are tap roots for Anselm Strauss's social worlds/arenas theory, which is foundational. Here, multiple collective actors—social worlds—participate in all

kinds of negotiations in broad and often contentious substantive arenas. Arenas are focused on matters on which all the involved social worlds and actors care enough to be committed to act and to produce discourses.

Second, SA draws upon the work of Michel Foucault, who challenged the social sciences by decentering the "knowing subject" (the individual human as agentic social actor) to focus on "the social" as constituted through discursive practices and on discourses as constitutive of subjectivities. Foucault asserts that framings, representations, linguistic conventions, and usages together constitute specific discursive fields or formations. Dominant discourses are reinforced through institutional systems of law, media, medicine, education, and so on. SA goes beyond "the knowing subject" to also analyze the salient discourses dwelling within situations. We are constantly awash in seas of discourse constitutive of life itself as well as individual and collective identities. SA enrolls poststructural approaches and follows in Foucault's footsteps to include analysis of historical, narrative, documentary, and visual discourses.

Third, SA requires inclusion of the *nonhuman,* underscoring the importance of understanding "things" in analyses of situations. SA takes the nonhuman elements in the situation of inquiry into account both materially and discursively. "Seeing" the agency of the nonhuman elements disrupts the taken-for-granted, demonstrating a materialist constructionism. Explicitly including the nonhuman in research also takes up the postmodern challenge of post-humanism—the idea that only humans "really" matter or "matter the most." Nonhuman elements may include cultural objects, technologies, animals, media, animate and inanimate pieces of material culture, and discourses—from cups and saucers to lab animals to TV programs. Nonhuman elements structurally condition the interactions within the situation through their specific material properties and requirements.

Three Kinds of Analytic Maps

In SA, the situation of inquiry is empirically constructed through making three kinds of maps, followed by analytic work and memo-ing in the GT tradition. The three very different kinds of "maps" are tools to provoke intensive analytical work. The first maps are *situational maps* that lay out the major human, nonhuman, discursive, historical, symbolic, cultural, political, temporal, and other elements in the research situation of focus. The goals of these maps are first to enhance research design by specifying everything about which at least some data should be gathered. Downstream, situational maps are used to provoke analysis of relations among the different elements. Working *against* simplification in postmodern ways, these maps capture and provoke analysis of the heterogeneous elements and messy complexities of the situation. Taking the nonhuman elements in the situation very seriously, materialities and discourses are also made visible in the situational maps.

Second, the *social worlds/arenas maps* lay out the collective actors and their arena(s) of commitment where they are engaged in ongoing discourse and negotiations. Such maps offer meso-level interpretations of the situation, taking up its social organizational, institutional, and discursive dimensions. They are distinctively postmodern in their assumptions: We cannot assume directionalities of influence; boundaries are open and porous; negotiations are fluid; discourses are multiple. Social worlds are universes of discourse, routinely producing discourses about elements of concern and other social worlds in the situation.

Third, *positional maps* lay out the major positions taken, and *not* taken, in particular discourses in the data. These are organized along specific axes of variation and difference, focus, and controversy found in the situation. Positions on such maps are *not* articulated with persons or groups but rather represent the full range of discursive positions taken on key issues in the situation. They allow multiple positions and even contradictions to be articulated. Narrative, visual, and historical discourses can all be analyzed.

In sum, SA offers three kinds of maps as fresh analytic devices to supplement grounded theory research. In centering on the situation, SA differs from the GT, with its focus on social processes—human action. Anselm Strauss's social worlds and arenas theory, Foucault's emphases on discourses, and the analytic centrality of the nonhuman are central to this poststructural approach to qualitative research.

Adele E. Clarke, Carrie Friese, and Rachel Washburn

See also Abduction and Inference to the Best Explanation; Discourse Analysis; Foucault's Thought; Pragmatism and the Social Sciences; Social Constructivism; Structuralism and Poststructuralism; Symbolic Interactionism

Further Readings

The Anselm L. Strauss Fund, UCSF Foundation. (n.d.). Anselm Strauss. Retrieved from http://sbs.ucsf.edu/medsoc/anselmstrauss

Clarke, A. (2005). *Situational analysis: Grounded theory after the postmodern turn*. Thousand Oaks, CA: Sage. (German translation, 2012, Wiesbaden, Germany: VS-Verlag für Sozialwissenschaften)

Clarke, A. E., & Friese, C. (2007). Situational analysis: Going beyond traditional grounded theory. In K. Charmaz & A. Bryant (Eds.), *Handbook of grounded theory* (pp. 694–743). London, England: Sage.

Clarke, A. E., & Star, S. L. (2007). Social worlds/arenas as a theory-methods package. In E. Hackett, O. Amsterdamska, M. Lynch, & J. Wacjman (Eds.), *Handbook of science and technology studies* (pp. 113–137). Cambridge: MIT Press.

Situational analysis. (n.d.). Retrieved from http://www.situationalanalysis.com

Strauss, A. (1993). *Continual permutation of action*. New York, NY: Aldine de Gruyter.

Situational Logic

This entry introduces the notion of situational logic or logic of situation in its most prominent form and explains its importance in social-scientific explanation of action and agency.

Situational analysis is a simplification of Max Weber's claim that one kind of social action—"instrumental rationality"—could be explained by the aims, means, and expectations of the actors. The improved formulation is found in Karl R. Popper's classic works *The Poverty of Historicism* and *The Open Society and Its Enemies*. It is not to be confused with W. I. Thomas's "definition of the situation," although it is easy to see the latter as a part of the former.

Situational analysis is the schema with the fewest assumptions for explaining the outcome of actions of agents and/or agencies. The social situation is analyzed into agents and agencies (usually social institutions), which have aims or directions; they are enclosed by and form part of an environment both physical and social; and agents and agencies possess putative knowledge of the situation and the leeway it allows them. For example, a driver setting out on a journey will have putative self-knowledge as well as putative knowledge of the destination, the possible routes, the condition of the vehicle, the roads, the traffic conditions, the applicable laws, and the weather. Since the vast majority of journeys are completed, we can say that most of the time the agents' appreciation of their situation is sufficiently congruent with the facts to enable them to reach their destinations. This should suffice for the explanation of the driver's conduct that ends with reaching the intended destination.

Situational Explanation

Agents and agencies that successfully accomplish their aims—reaching a destination, collecting taxes, prosecuting a military engagement—pose explanatory problems that are trivially soluble. Failures and unintended and counterintended outcomes require explanations that are much less obvious. Researchers seek such explanations by conjecturing different aims, situational conditions, and actor and institutional knowledge, to see if they will explain the outcome. If such conjectures are promising as explanations, then they deserve empirical testing.

Social outcomes that no agent or agency intends intrigue social scientists: for example, suicide rates and divorce rates; actors who seemingly act against their interests in eating poisons, purchasing rubbish, voting, and so on. We can bring out the structure of situational explanations of the unintended by analyzing accidents, intended by no one, and especially catastrophic accidents, where the outcome was not even imagined.

Here is such an example. In November 1987, a fire in a small escalator at King's Cross Station on London's Underground transport system grew rapidly into a "flashover" that engulfed a ticket hall in smoke and flames, killing 31 people and injuring and maiming many others. The official investigation into the fire did not suggest that the lighted match that ignited it was intended to do so. More than likely, a smoker was "lighting up" on the way out of the nonsmoking system. The flammable mixture of dirt, grease, and fluff under the (wooden) escalator steps

was due to sloppy cleaning protocols (an unnoticed institutional deficiency). The lack of a central control room, of fully trained staff, and of working communications systems were due to neglect of the physical and social systems involved, not by design. The neglect had seemed unproblematic, since several such fires were extinguished every week and none had been fatal. While scientists and engineers theorized that the sloping tunnels caused the flames to shoot upward in search of oxygen and hence accelerate beyond all human capacity to cope, the investigation also looked at the many physical, social, and cultural configurations of the Underground system and pinpointed faults in each one that contributed to the sequence of events that constituted the disaster.

The King's Cross fire was a complex situation analyzed to find its various logics: the physical logic of fire in upward-sloping tunnels; the psychosocial logic of smokers dying for a "gasper" as they approach the surface; the social logic of lack of training, central direction, and clear priorities for members of staff; the logic of station design that showed up poorly in interrogations about fire control and evacuation; the logic of actions by staff, police, firemen, and emergency responders unfamiliar with the layout of the labyrinth of tunnels and hence of the best means of egress; the logic of the expectations of all concerned, namely, that small fires pose little threat: the result—no announcements, no fire alarm, no sprinklers activated.

The aim of passengers is to complete their journey. The aim of the divisions of the London Underground railway is to maintain its smooth, safe, and economical functioning. Of necessity, the knowledge of all participants is partial and often flawed; hence, they did not assess the developing situation correctly and acted in ways that were always suboptimal but usually not sufficiently below standard to be dangerous, until the flashover accident. Especially, all parties were unaware of the possibility of a flashover from a small fire and hence were unable to take that into account in their actions.

Critical Objections

This simple and intuitive model faces criticisms from holists, who see its individualism as superficial, and strict methodological individualists, who see it as not individualist enough. Leaving aside the former objection as transcendental, the latter stipulates that agencies such as groups (the passengers and the station staff) and institutions (the police, the fire services) do not have aims. In fact, we often speak as though such wholes do have aims—the way we often speak even of inanimate things (see the statement above about flames shooting upward in search of oxygen). Mostly, such locutions can be treated as shorthand. Institutional and group actions can be further analyzed into the actions of particular individuals playing their institutional roles, carrying out policies or instructions, or into aggregates of individual actions, for example, votes. The aggregation view of institutional actions has been criticized as being unable to account for shared or collective intentions (e.g., "We intend to have two children"), as with link couples, policemen ("We intend to get to the bottom of this"), and revelers ("We shall party"). Even more severe is the objection that situational models are static in treating the agent, the aims, and the agent's situation as fixed while the actions and their reverberations play out, when in fact, as is obvious in all three examples, they may change (two children may become three, cases go "cold," revelers get bored). John Wettersten argues that the agent's thoughts and, hence, aims are formulated within institutional constraints, institutions that are in constant flux, not the least because of the agent's thinking within and through them. This explains, he says, why Weberian models are nonplussed by social change—even social change initiated by the agents. As a result of these caveats, care needs to be taken with situational explanations not to reify institutions, not to make them into actors as opposed to enablers and constraints, and not to freeze the flux of developing situations.

Ian Jarvie

See also Action, Philosophical Theory of; Explanation, Theories of; Holism, in the Social Sciences; Individualism, Methodological; Institutionalism and Institutional Theory; Popper's Philosophy of Science; Rationality and Social Explanation; Weber's *Verstehende* Approach

Further Readings

Fennell, D. (1988). *Investigation into the King's Cross underground fire.* London, England: HMSO.

Popper, K. R. (1945). *The open society and its enemies.* London, England: Routledge.

Popper, K. R. (1957). *The poverty of historicism.* London, England: Routledge. (Reprinted from The poverty of

historicism I, II, & III. *Economica, 11,* 86–103, 119–137 and *12,* 69–89; 1943–1944)
Thomas, W. I. (1923). The definition of the situation. In *The unadjusted girl.* Boston, MA: Little, Brown. Retrieved from http://www.sociosite.net/topics/texts/thomas.php
Weber, M. (1968). *Economy and society: An outline of interpretive sociology.* Berkeley: University of California Press. (Original work published 1922)
Wettersten, J. (2006). *How institutions steer events.* Aldershot, England: Ashgate.

Social Anthropology

Social anthropology focuses on the social, political, religious, and economic organization of human groups. It has traditionally understood these as aspects of culture, and it often emphasized the symbolic and meaningful dimensions of human activity. Within the discipline of anthropology, *social* anthropology is distinguished from physical anthropology, which studies biological variation among humans and our evolutionary history. In the United States, the term *cultural anthropology* is used to refer to the same range of concerns.

Throughout its history, social anthropology has both drawn on and contributed to philosophical reflection. In the 19th century, social anthropologists were engaged in a theoretical enterprise that was closely related to philosophical writing about human nature. Twentieth-century philosophy of language and philosophy of mind have been important sources of inspiration for social anthropologists. Philosophers have also debated the consequences of anthropological theories and empirical findings. Social anthropology has thus been an important resource for philosophical work on rationality, ethics, language, politics, the mind, race, gender, and social ontology. Finally, social anthropology has been at the forefront of the development of empirical methods for the study of humans, and it has therefore been a party to debates about the unity of science and the character of scientific knowledge.

The Philosophical Roots of Social Anthropology

Social anthropology in the 19th century had a broad historical scope. Global trade and colonization since the 16th century had made European intellectuals aware of the wide variety of human societies. Philosophers had long commented on these differences and often tried to account for them by imagining the developmental stages through which "primitive" humans became "civilized." Philosophical speculation on this topic was primarily aimed at demonstrating how human societies could, in principle, have arisen through the action of fundamental principles of human nature. The 19th-century social theorists developed these speculations into an empirical science. Lewis Henry Morgan, for instance, based his theory of social evolution on long-term interaction with the American indigenous peoples. Edward Burnett Tylor collected comprehensive information from missionaries, travelers, and colonial officials. The 19th-century anthropologists explained observed differences by treating contemporary peoples as exhibiting features of earlier stages of human development. Like their philosophical colleagues, they tended to see the evolution of more complex societies as reflecting either the development of rational thought or the victory of rational thought over superstition.

Romanticism, a broad literary, artistic, and philosophical movement, ran counter to the more rationalistic evolutionary narratives of the 19th century. Romanticism valorized folkways, and it thought of humans as naturally divided into nations or peoples. Johann Herder argued that language determines thought, and he postulated a very close relationship between languages and nations. Philosophy in the romantic spirit thus provided support for the later development of the concept of *culture.*

The Unity of Science

As the discipline of social anthropology became more empirical, it began to develop new methods. By the early 20th century, extensive fieldwork (ethnography) was regarded as necessary to any description of a culture. Ethnography relied on the method of participant observation. The anthropologist lived among his or her subjects for an extended period (typically a year or more), learning their language, participating in their daily routines, and conducting extended interviews. These methods are different from those found in the natural sciences, and they engage the question of the unity of science: Are different forms of knowledge associated with the study of human subjects and nonhuman objects?

Social anthropologists and philosophers have taken up both sides of this debate. Many have argued that anthropology is hermeneutic. It articulates the meaning of symbols, actions, rituals, and life experiences. Such meanings are always particular and, hence, not subject to generalization into laws of psychology, economics, or biology. Others have responded that emphasis on interpretation to the exclusion of causal explanation limits the interest and ambition of the field. Some, influenced by Karl Marx's materialism, have gone so far as to argue that the cultural meanings are nothing more than ideologies that hide the real mechanisms of social life. A more modest view is that hermeneutic and causal/explanatory inquiries are complementary. In practice, social anthropologists tend to be methodologically pluralistic. The philosophical challenge of such pluralism is to explain how specific methods support each other and how conflict among them is to be resolved.

Functionalism was an important theoretical position within social anthropology in the first half of the 20th century, and it is relevant to the debate over the unity of science. Functionalists explained the existence of specific social practices in terms of the benefits that the practice had for society. Accusations of witchcraft, for instance, were explained by their function in resolving social conflict; the accusation, investigation, and retribution (if any) diffused social tensions. Prima facie, functional explanations seem distinct from other causal explanations. Where a causal explanation appeals to a temporally prior event to explain a later event, functional explanations are "teleological," in the sense that they rely on goals or future goods. Philosophers of science have debated whether functional explanation is a legitimate form of scientific explanation and, if so, whether it can be assimilated to a general theory of explanation. While functionalism is largely out of fashion among social anthropologists today, it was an important element of 20th-century debates about the unity of science.

Culture

In its 19th-century form, the concept of *culture* retained its connotations of learning and high art. Social anthropologists used "culture" and "civilization" as synonyms and thought in terms of culture evolving from lower to higher forms. The 20th-century anthropologists rejected these hierarchical ideas and thought of the human social world as populated by many distinct cultures. A culture was an integrated system of ideas, norms, values, practices, decorative motifs, material objects, and so on, passed from one generation to the next. This commitment to "holism" entailed that the interpretation of any single item required reference to other items. It also entailed that cultures were relatively independent, both ontologically and historically. While there were historical relations (of trade, etc.) among cultures, each culture needed to be understood in its own terms. The idea of culture as something (a) distinctive of and shared by a particular group of people, (b) not reducible to either biology or psychology, and (c) constituted, at least in part, by holistically related meanings, symbols, and values has been profoundly influential both inside and outside the academy.

The concept of culture came under sharp attack in the late 20th century. A mix of political and epistemic arguments concluded that the portrayal of a given culture was an artifact of anthropological writing, not an object that existed prior to anthropological inquiry. In addition, anthropologists became dubious of the idea that cultures were stable and shared. Indeed, some began to doubt the analytic usefulness of the concept of culture at all.

There have been three broad responses within social anthropology. The first is sometimes identified as *postmodern*. These anthropologists tend to see human groups as fragmented, disunified, and driven by power. They often experiment with different ways of giving voice to their interlocutors. A second response is known as *practice theory*. This view dissolves cultures into multiple practices that overlap and conflict. Drawing on Ludwig Wittgenstein's work on rule following, Pierre Bourdieu conceived of practices as normative ways of acting that are passed among individuals. Practice theory in anthropology is related to the recent use of the concept of practice in philosophical accounts of normativity. In an important critique, Stephen Turner argued that practice theory cannot provide a coherent account of its central idea: practice.

The third response to the demise of the culture concept has been to look to the cognitive grounds for social phenomena. These writers treat a culture as a population of individuals who have more or less similar representations. Cognitive mechanisms for the learning, storage, retrieval, and transmission of representations explain the persistence of some ideas and the change in others. This perspective

has permitted the integration of social anthropology with cognitive psychology and has suggested provocative explanations of religious, political, economic, and other phenomena. Philosophically, this trend is interesting because it raises again the question of whether human phenomena can be explained in causal terms or whether other, more hermeneutic modes of understanding are required.

Rationality and Relativism

Social anthropology has sparked lively debates about the culturally relative character of human thought and value. Anthropological studies demonstrate substantial variation, but mere variation in avowed norms or concepts does not constitute relativism. A relativist must affirm both a descriptive thesis (that variation exists) and a thesis about the *incommensurability* of the norms (etc.) in question. Anthropological fieldwork seemed to demonstrate substantial human variation, and the concept of culture espoused by midcentury anthropologists provided a theoretical background for an incommensurability argument. The content of all norms (including our own) depends on a culture's religious, economic, political, or social institutions, and hence, there were no a-cultural (absolute, objective) norms by which to adjudicate differences in opinion. Insofar as the debate about relativism presupposed the classic concept of culture, it has been somewhat undermined by the recent critique mentioned above.

Rationality and Irrationality in Interpretation

Social anthropologists such as James Frazer and Lucien Lévy-Bruhl contended that human thought evolved from less to more rational forms. Midcentury social anthropologists rejected evolutionary claims, but some kept the idea that rationality took culturally specific forms. The philosophical debate concerned whether variation in forms of reasoning was possible or empirically discoverable. Peter Winch argued that ways of thinking were deeply related to forms of life and that a good interpretation must be sensitive to such differences. Donald Davidson argued for a "principle of charity," which demanded that an interpreter attribute true beliefs to the subjects when their utterances are, in the opinion of the interpreter, true. The principle of charity entails that any interpretation must portray the subjects as conforming to standard canons of logical inference; hence, the possibility of alternative modes of thought imagined by Winch and some anthropologists is illusory. A middle ground of the debate held that while variation in reasoning might be found in esoteric areas, such as religion or myth, interpretation of these domains required agreement in mundane areas.

Moral Relativism

Social anthropologists have often espoused (or been accused of espousing) moral relativism. While the empirical studies of social anthropologists have been taken as supporting the descriptive thesis of moral relativism, there has been some debate about the true range of human moral variation. Some have argued that there is substantial agreement on the general moral principles concerning lying, adultery, theft, and murder and that the variation concerns the justifiable exceptions. Recent work in the cognitive foundations of morality has supported this position.

Both anthropologists and philosophers have argued that cross-cultural moral disputes cannot be rationally adjudicated. In these discussions, it is important to distinguish arguments for *tolerance* from arguments about incommensurability. It is often claimed that a proper understanding of other cultures requires the interpreter to refrain from evaluation. Strong versions of this claim have been disputed by pointing out the importance of evaluation in any scientific research. A more modest claim is that recognition of cultural differences can motivate tolerant attitudes. However, tolerance of moral disagreement does not entail that there is no way to rationally resolve differences; hence, it does not lead to moral relativism. Moreover, even this weak claim about tolerance has been controversial within anthropology. Anthropologists are often concerned with human rights or the political status of ethnic minorities and, therefore, have sought to make and justify cross-cultural moral judgments.

Mark Risjord

See also Causes Versus Reasons in Action Explanation; Cognitive Anthropology and Mental Architecture; Cultural Evolution; Economic Anthropology; Ethnography, Philosophical Aspects; Naturalism in Social Science; Relativism and the Social Sciences: From the Sapir-Whorf Hypothesis to Peter Winch; Relativisms and Their Ontologies; Social Practices; Structural Functionalism, in Social Theory; Symbolism

Further Readings

Evans-Pritchard, E. E. (1951). *Social anthropology.* London, England: Routledge.

Geertz, C. (1973). *The interpretation of cultures.* New York, NY: Basic Books.

Krausz, M. (Ed.). (1989). *Relativism: Interpretation and confrontation.* Notre Dame, IN: University of Notre Dame Press.

Ortner, S. (1984). Theory in anthropology since the sixties. *Comparative Studies in Society and History, 26,* 126–166.

Sperber, D. (1996). *Explaining culture.* Oxford, England: Blackwell.

Turner, S., & Risjord, M. (Eds.). (2006). *Philosophy of anthropology and sociology.* Amsterdam, Netherlands: Elsevier.

Winch, P. (1964). Understanding a primitive society. *American Philosophical Quarterly, 1,* 307–324.

Social Anti-Individualism and the Mental

This entry gives an overview of a powerful philosophical thesis holding that the concepts figuring in an individual's mental states are determined as to what they precisely mean not simply by what the individual believes or thinks she means by them but also by her linguistic community—that is, by the social environment. The entry distinguishes between two basic kinds of such mental anti-individualism and points out its implications for social sciences and, in particular, for one of them, psychology.

The Thesis

Social anti-individualism (also known as *social externalism*) is the thesis that relations between an individual and members of her linguistic community help determine which concepts she possesses and hence which thoughts she is capable of thinking. This means that, contrary to popular opinion, an individual's mental states are *not* determined entirely by her internal, physical states (e.g., her brain states); rather, they are determined by her internal, physical states *in conjunction with* her relations to an *external, social environment*.

The thesis is widely accepted in philosophy and has a number of important implications concerning, for example, the individuation of kinds in psychology, the nature of causation and explanation in psychology, and the relation between psychology and the natural sciences. Similar implications hold for the social sciences generally.

The *Twin Earth* Thought Experiments

Social anti-individualism is established by reflection on a certain kind of thought experiment (made famous by Tyler Burge and Hilary Putnam), an example of which runs as follows. First suppose that a subject, Sam, has a number of true beliefs about games—she believes that some games are more fun than others, that chess is a game, that children like party games, and so on. In addition to these true beliefs about games, she also believes falsely that games must involve at least two people. Further suppose that if her friends were to point out to her that patience and solitaire are games, she would accept their correction and update her use of the term *game* to match with her improved understanding of the term and to reflect her fuller understanding of the concept *GAME*.

Now consider a second scenario in which Sam's intrinsic physical states are the same while her linguistic community is different. In this second scenario, let us suppose, it is standard practice to use the term *game* as Sam does. That is, in the second scenario, the term *game* is defined to apply (roughly) to competitive, recreational activities involving at least two people. The first and second scenarios, then, have a word form in common ("game"), but the term has a different meaning in each, which is to say that it expresses a *different* concept in each. In the first scenario, the term *game* expresses the concept *GAME,* which correctly applies to activities including solitaire and patience (even though Sam does not initially realize this). In the second scenario, in contrast, the term *game* does not express the concept *GAME* but rather expresses a different concept (albeit a similar one—let's call it *SPAME*) that does not correctly apply to activities such as solitaire and patience. Consequently, while Sam possesses the concept *GAME* in the first scenario, she does *not* in the second, even though her physical states and classificatory practices are *identical* in both scenarios. According to the social anti-individualist, then, social differences—specifically differences in the way in which people around us use words, that is, our linguistic communities—can affect our mental states.

Natural-Kind Anti-Individualism

The thought experiment relies on a number of assumptions: that words express concepts; that understanding the meaning of a word is a matter of degree and, hence, that understanding a concept is a matter of degree; and that the meaning of a word is determined by the way in which that word is used by the community as a whole, rather than by the way in which it is used by any given individual. Although these assumptions are plausible, there is a different kind of anti-individualism—*natural-kind* anti-individualism—that does not rely on them.

According to natural-kind anti-individualism, it is relations between an individual and her natural environment that play a role in determining which concepts she possesses and, hence, which thoughts she is capable of thinking. Specifically, natural-kind concepts, such as the concepts *WATER, SILVER, TIGER, OAK,* and so on, are thought to be individuated in part by objective relations between the thinker and instances of those natural kinds in her environment. If an individual is related in the right kind of way to water, then she may be able to think about water by means of the concept *WATER*. If, on the other hand, an individual bears no relations to water at all (perhaps because she lives in a world with no water), then she will be unable to think about water by means of the concept *WATER,* for how could she have acquired that concept? Natural-kind anti-individualism is the more widely accepted form of anti-individualism, but social anti-individualism is the more general thesis, covering all concepts rather than merely natural-kind concepts.

The Precise Target of Anti-Individualism

It is crucial to understand anti-individualism as a thesis about the individuation conditions of concepts rather than as a thesis about the causes of particular thoughts involving those concepts. Thus, the core claim of anti-individualism is that relations between an individual and her environment play an essential role in determining not just what thoughts that individual happens to have at a particular time but the very concepts that enable her to represent the world in thought at all. Interestingly, both forms of anti-individualism have the same implications for psychology, as follows.

Implications for Psychology and the Social Sciences

First, anti-individualism rules out all forms of physicalism: Two individuals can have the same physical properties and yet have different mental properties. Second, this means that psychology is not reducible to physics. Third, this in turn means that psychological causation between psychological kinds and psychological explanation that invokes psychological kinds are to be treated as in a certain sense independent of physical causation and explanation that invoke physical kinds.

Similar implications hold for the social sciences generally since the philosophical thesis of mental anti-individualism shows that facts about an individual (in particular facts about her mental states and therefore the propositional attitudes figuring in them) cannot be construed solely as facts about that individual alone; facts *external* to the individual involving social facts about linguistic use, and so on, enter into the individuation of the concepts she uses. Hence, a social-scientific explanation of what an individual believes or not must take these social facts into account, going "beyond the individual" as such.

Sarah Sawyer

See also Holism, in the Philosophy of Language; Holism, in the Social Sciences; Individualism, Methodological; Kinds: Natural Kinds Versus Human Kinds; Mind–Body Relation; Supervenience; Thought Experiments

Further Readings

Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy, 4,* 73–121.

Burge, T. (1986). Intellectual norms and foundations of mind. *Journal of Philosophy, 83*(12), 697–720.

Burge, T. (1993). Mind-body causation and explanatory practice. In J. Heil & A. Mele (Eds.), *Mental causation* (pp. 97–120). Oxford, England: Oxford University Press.

Putnam, H. (1979). The meaning of "meaning." In *Philosophical papers: Vol. 2. Mind, language and reality* (2nd ed., pp. 215–271). Cambridge, England: Cambridge University Press.

Sawyer, S. (2011). Internalism and externalism in mind. In J. Garvey (Ed.), *The Continuum companion to the philosophy of mind* (pp. 133–150). London, England: Continuum.

Social Capital

Social capital (SC) may be understood as the benefit or utility that humans may obtain by entering into ongoing social relations or social structures. This entry presents the conceptual, methodological, and critical issues characterizing the importance of SC for the social sciences.

As is often the case in the social sciences, in SC, too, there is no agreement on what the concept encompasses. While SC is perceived in divergent ways, with a plurality of approaches and empirical operationalizations being available, yet there is little discussion, let alone agreement, among dissenting viewpoints or perspectives. The postmodernist practice of allowing or elevating to equal status all variant interpretations has perplexed the issue even further. Insistence on the true meaning of the term risks falling into a kind of essentialism, something that is denounced, even if the reverse side of this appears to be methodological and conceptual nihilism. SC has thus been defined in ways that complement or conflict one another. It appears that such a definitional variance in SC is related to disciplinary-specific uses of the notion, varying from one social science to another. Such a prominent divergence among social sciences is the one separating the uses of SC in sociology and political science. Still, definitional variance is often found within a single discipline, too.

That being said, a general criticism of all the various versions of the SC notion questions its very essence, namely, its being construed as "capital." The "capital" in SC is considered by critics a travesty of the original and standard notion of "economic capital." Unlike "economic capital," which is transparent, specific, and clear, SC appears to some as meaningless or nonexistent. SC is seen as only a metaphor, and according to such critiques, it is void of meaning beyond that level.

Perspectives From Sociology

Pierre Bourdieu and James S. Coleman are the key sociologists exploring SC, albeit each perceives it differently. Bourdieu conceptualizes SC in close connection to other forms of capital, specifically the economic and the cultural varieties. The family is seen as a basic source of SC, and it is found exclusively among the socially powerful, namely, in the upper middle class or haute bourgeoisie. This happens because Bourdieu defines social class in terms of the possession or lack of capital, which assumes the aforementioned three forms (economic, cultural, and social); the dominated social classes and strata do not possess capital at all, including SC. He designates SC as comprising social responsibilities, "connections," or "linkages," and under certain circumstances, it may be converted into economic capital. For Bourdieu, SC is formed, more or less consciously, through one's integration into networks, and unlike economic capital, it has no specific material form, nor is it transparent. Instead, it is characterized by indeterminacy, so that there can be, for example, a leftover sense of an unspecified obligation. This, according to Bourdieu, is an inevitable dimension of SC. Otherwise, if it were clear and specific, it would simply be a series of ordinary nonmarket transactions. This conceptualization contrasts with other approaches to SC, particularly those suggested by Robert D. Putnam (see the next section).

On his part, James S. Coleman introduces the notion of SC as a "means of support," in the specific sense of enhancing students' performance, which will strengthen the generation of human capital—that is, of knowledge. Therefore, he aligns with Bourdieu in perceiving or using SC in connection to other, more typical forms of capital. For Coleman, SC results from the changes that take place among individuals, changes perceived to facilitate social action. Coleman thus defines SC on the basis of its function. Moreover, it is also determined by the outcome. Besides, Coleman, in agreement with Bourdieu, stresses the nonconcrete, nonmaterial, and indefinite character of SC when compared with other, more typical forms of capital. Furthermore, Coleman notes that SC is a public good and that this feature characterizes it as well as differentiates it from other forms of capital. It is seen to result from the empirical fact that the subject of generation of SC enjoys only a limited part of its overall benefits. It also implies that SC is not just or solely a property of, or benefit accrued to, the isolated individual agent who generates it, but it is also something enjoyed by other individuals as well as by the community. Furthermore, because SC operates within concrete social contexts, certain characteristics of the social relations can either facilitate its appearance

or, by contrast, impede it. Such characteristics are the availability of trust and reciprocity among the members of the inner group, as well as the effectiveness of existing normative regulations. Overall, Coleman's application of SC has been in linking the generation of educational achievement to, primarily, micro-social-level influences. Yet his rather unclear definition of SC has, arguably, opened up the way for the application of the SC label to a range of various and contradictory procedures.

Putnam's Views and Critique

In the area of political science, the key figure in SC studies is Robert D. Putnam, who explicitly draws on Coleman's characterization of the concept. However, on Putnam's conception, SC and trust are explicitly linked to each other. SC is generated by linking social trust, social norms, and social networks within social organizations, enhancing the efficiency of socially coordinated actions.

In Putnam's subsequent elaborations, there is transference of meaning; the radius of the SC notion expands from the meso level of analysis to the macro level. Thus, it comes to be formed by "features of social life." This broadening contrasts with the individual level of the actors (personal or collective, as Bourdieu and Coleman had it). What is more, the broadening of the SC notion itself has been neither explained nor justified.

On Putnam's conceptualization of SC, any differences in economic, social, or other forms of power do not raise a significant issue. What, however, is the significant element is the participation of individuals in the whole gamut of possible social groups and collective activities available in any given social context. In other words, the key feature is participation per se, as well as the extent to which it appears—participation being the underlying assumption and hence of major importance in political science. In fact, in this neopluralistic approach, a participatory attitude and activation within the context of community networks seems to generate additional forms of SC. Thus, SC can do, as well as accrue from, the "bonding," "bridging," or even "linking" of social groups. In fact, in Putnam's approach, the SC "stock" equals the degree of participation in various social activities and organizations, at all levels, such as in political organizations, scientific associations, or neighborhood choirs.

Putnam's construal of SC has attracted several critiques—some of which echo the social-scientific disciplinarity divergence between political science and sociology mentioned at the start. "Participation" is claimed as requiring to be concretized and suitably specified since not every kind of participation is beneficial; besides, the scale of groups involved is most important, and issues of size must be factored in when measuring SC. It is claimed that Putnam fails to offer a substantial distinction between various kinds of participation, so critics working in sociology reject his use of it as outright meaningless. Another critique concerns the issue of the logical circularity claimed to endanger Putnam's claims about SC, as well as the indeterminacy of causal direction: in other words, that cause and effect are entirely blurred and their order circular, rendering such an understanding of what SC is nondemonstrable. Furthermore, there are serious objections to the relevance, precision, and sufficiency of the indices used to empirically ascertain SC. Particularly, the idea of "density of relations" between the members of a network, measured, for instance, by the frequency of members' contacts, as favoring the generation of SC, as well as providing proof of its existence, has met with difficulties. Furthermore, SC, critics argue, must not be assumed to be an inherently positive social condition; it may be negative too, as in "mafia"-type criminal networks that promote members' welfare at the expense of outsiders or as in the case of groups adopting exclusionary practices beneficial to them, such as the "closed shop."

What, however, should not be ignored is that Putnam links SC to trust, and he does indeed have a notion of trust. He explicitly states that trust "is a form of SC," and indeed "an essential component of SC," and holds that trust is generated by norms of reciprocity and networks of civic engagement through socialization, an approach that relates to his being a political scientist. However, it is evident that for Putnam, largely because of the problem of causal indeterminacy (see above), SC and trust end up as near-identical notions. This is further conflated in empirical applications when trust is taken as a proxy for SC.

From a theoretical point of view, this practice is quite problematic because it disregards the sociological tradition indicated above, according to which SC is basically about in-group members accessing various resources. This process of ensuring resources is

assisted by the prevalence of trusting relations. This means that having trust relationships is a useful or even necessary, but by no means a sufficient, condition for possessing SC. Accordingly, using the latter as a proxy for the former will not do.

Nevertheless, the influence of Putnam's conception of SC has been quite significant in all social sciences, and it is to be found within sociology too. This is probably for no good analytical reason but merely for "convenience sake" as it facilitates the provisioning of findings in figures that are dear to political elites, which have been utilizing the SC notion for their own purposes almost from its inception. In this sense, this particular conception of SC operates ideologically, as Alejandro Portes has pointed out.

Alejandro Portes: An Economic Sociology Approach

For the economic sociologist Alejandro Portes, to possess SC, one must relate to others who constitute the source of SC and of the privileges it provides. At this point, however, as he notes, there appears to be confusion in the literature as the source or origin of the SC is repeatedly and falsely regarded as an outcome of the action of SC. Portes has observed this confusion in both Coleman's and Putnam's work, as well as in other writers relying on them. But in this way, causality is not defined; it becomes a vicious cycle, and eventually tautology prevails.

To deal with this problem, Portes and his colleagues urge for a clear distinction between the sources or origin, in other words, the cause of the SC, and the results or effects of its action. Thus, four sources from which SC originates are identified: (1) internalization of values, (2) transactions of a reciprocal character, (3) forms of collective solidarity, together with (4) the trust imposed by negative or positive sanctions. Besides, the sources of SC are embedded in the motives the members of a network or a group have in order to provide resources. These motives can be distinguished, on the one hand, into consummatory ones, which may derive from the initial socialization (e.g., within the family) and into the consecutive internalization of certain values or regulatory patterns, or may be cultivated, in the context of the community (hence marked by bounded rationality, i.e., by decisions that are rational given the specific constraints under which they are taken, as in the case of solidarity strikes). On the other hand, they may originate from instrumental motives, in which there is expectancy of reciprocity and trust, as when the sponsor is secured against fraud. Thus, the various sources of origin of the SC lead to its composite formation, so that SC is perceived as the ability to secure benefits via participation in networks and other social structures.

On the whole, Portes's interventions offer a more balanced understanding of SC and its potential. The notion is not rejected but specified and rationalized sociologically. The emphasis is on the need to systematically study the precise effects of SC and avoid attributing irrelevant, accidental, or spurious effects to it. This perception has led to conceptions of SC more akin to the micro level that focus on the individual's relationships to her or his network of social connections and the benefits and resources he or she may muster.

Sokratis Koniordos

See also Economic Sociology; Public Goods; Social Networks; Trust, Social

Further Readings

Bourdieu, P. (2001). Forms of capital. In M. Granovetter & R. Swedberg (Eds.), *The sociology of economic life* (2nd ed., pp. 96–111). Boulder, CO: Westview Press.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*(Suppl.), S95–S120.

Koniordos, S. M. (2008). Social capital contested. *International Review of Sociology, 18*(2), 317–337.

Mouw, T. (2005). Social capital and finding a job: Do contacts matter? *American Sociological Review, 68*, 868–898.

Portes, A. (1998). Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology, 24*, 1–24.

Portes, A. (2010). *Economic sociology: A systematic inquiry.* Princeton, NJ: Princeton University Press.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community.* New York, NY: Touchstone.

Putnam, R. D. (with Leonardi, R., & Nanetti, R. Y.). (1993). *Making democracy work: Civic traditions in modern Italy.* Princeton, NJ: Princeton University Press.

Torche, F., & Valenzuela, E. (2011). Trust and reciprocity: A theoretical distinction of the sources of social capital. *European Journal of Social Theory, 14*(2), 181–198.

Social Choice Theory

The theory of social choice considers the problem of aggregating the preferences of the members of a given society in order to derive a social preference that represents this society or community. This entry reviews the major aspects of Social Choice Theory and explains the main theories and advances in it.

The social preference is to express the general will, the common good as it were. The general will can be viewed as the basis for the very existence of any society. Economists argue that the common good finds its expression in a so-called social welfare function, which, described more mundanely, represents a compromise among the divergent interests of those who belong to society. The market mechanism, in contrast, cannot be taken as a social welfare function, since it is not guided by moral or ethical principles in any deeper sense. A market allocation heavily depends on the initial endowments of the individuals. These possessions determine the power or weakness of the individual agents.

Influences on the development of Social Choice Theory have been manifold over the centuries. Mathematicians, social scientists, and philosophers made important contributions of different kinds. The Marquis de Condorcet (1743–1794) and Jean-Charles de Borda (1733–1799) explored the majority rule and ranking methods at the time of the French Revolution. Roughly in the same period, Adam Smith (1723–1790) elaborated the concept of an impartial observer. The utilitarian philosophy of Francis Hutcheson (1694–1746), Jeremy Bentham (1748–1832), and others has been most influential over several centuries, with its modern-day version brought forward by John Harsanyi (1920–2000) in the middle of the past century. John Rawls's (1921–2002) theory of justice, from around 1970, became a powerful contestant of utilitarianism over the past few decades. It is probably fair to say that the modern theory of collective choice started with Kenneth Arrow's (1921–) pathbreaking work on the nonexistence of a social welfare function around 1950. All these works and several more will be discussed in what follows.

Impossibilities

If a social welfare function is an expression of the general will of the populace, it should be able to deal with whatever kind of preferences the individual members of a given society have. More technically speaking, and this is Arrow's definition, a social welfare function is a mapping from the set of all logically possible combinations of individual preference relations over a given set of social states or alternatives to the set of all logically possible orderings over these states. This is the requirement of *unrestricted domain*. In other words, it should not be admissible that any individual ranking be a priori excluded, for whatever reason, from the set of preferences of the members of society. The next requirement, called *the weak Pareto principle,* says that if for any two alternatives x and y, say, all members of society agree that x, for example, is strictly preferred to y, then society should have exactly the same strict preference. Next, information gathering for the aggregation procedure should be parsimonious, that is, if society has to make a decision between two alternatives (let's call them again x and y), the individual preferences with respect to x and y only, and not any other preferences, should be taken into account in order to distil the social ranking between these two alternatives. Arrow calls this condition the requirement of *independence of irrelevant alternatives*. Last, we do not want that there exist a particular person in society such that whenever this person has a strict preference for some alternative over another, society "automatically" has the same strict preference, for any two alternatives and any preference profile of the members of society. This is called the *non-dictatorship* condition. For Arrow, these four conditions are necessary requirements for a democratic decision procedure—perhaps not sufficient since there may be other demands as well. Unfortunately, in the case of at least three alternatives, these four conditions cannot be simultaneously fulfilled by any social welfare function. In other words, there does not exist a social welfare function satisfying these four requirements. This is Arrow's famous *impossibility theorem*.

Clearly, almost everyone will agree that there should not exist a dictator in society who via his or her own strict preference over any pair of social alternatives automatically determines the social preference over these alternatives. Arrow defined social alternatives as very complex social states, which can include, among other things, the issue of waging a war against another nation or introducing the death penalty for certain crimes. On the other hand, democracies should allow individuals a certain

amount of freedom and autonomy over purely private matters. John Stuart Mill, in the 19th century, spoke about a circle around every human being that nobody should be allowed to intrude into. So "local decisiveness" to a certain degree should be permissible. In our time, Amartya Sen (1933–) was the first to integrate this idea into Arrovian social choice. He formulated that each and every individual be permitted to exercise local decisiveness over at least one pair of social states, which differ only with respect to some private matters of that particular person. He then proved another important impossibility result, "the impossibility of a Paretian liberal." He showed that the requirements of unrestricted domain and the weak Pareto principle are incompatible with the exercise of local dictatorship or individual autonomy over purely private matters. In other words, there does not exist a social welfare function fulfilling these three conditions.

Wulf Gaertner, Prasanta Pattanaik, and Kotaro Suzumura, among others, have argued that the Arrovian setup may not have the appropriate structure to formulate the exercise of personal rights; the game form structure within which individuals select certain actions that shape private features may be closer to what one observes in real life, where closeness refers to conformity with existing rights systems. Notice, however, that this alternative suggestion by no means denies that there can exist a clash between the exercise of individual rights and the Pareto principle. A reference to findings in noncooperative game theory may be adequate here, namely, the existence of Nash equilibria that are Pareto dominated.

A third important impossibility theorem in Social Choice Theory refers to the fact that it can be advantageous for individuals to misrepresent their preferences and, by doing so, to achieve an outcome that is more favorable for them than the one that would have come about if they had announced their true or honest preferences. This phenomenon is well-known in the allocation of public goods, where individuals may want to hide their true willingness to pay in order to achieve a lower contribution fee for themselves. In Social Choice Theory, Allan Gibbard, a philosopher, and Mark Satterthwaite, an economist, were the first to independently provide a deeper analysis of this phenomenon. Of course, there are social choice or aggregation rules that are not manipulable, but these are highly unsatisfactory in the sense that they do not respond at all or respond only very faintly to changes in individual preferences. The ranking rule proposed by the French mathematician we mentioned above, Jean-Charles de Borda, which we shall discuss a little later in this entry, is highly susceptible to manipulation (Borda was very well aware of this fact, so he declared that his method was only for honest men), since its aggregation scheme employs very detailed information coming from the individual rank orderings. Gibbard and Satterthwaite proved that if individual rank orderings are unrestricted, if furthermore the aggregation method is "monotonic" or responsive to changes in the individuals' preference rankings and Pareto efficient, and if there are at least three social alternatives, the only nonmanipulable or "strategy-proof" aggregation method is dictatorial.

Though there are lots of other negative results in Social Choice Theory, there is wide agreement among researchers in this area that Arrow's, Sen's, and the last result about strategy-proofness are the most important. We shall now turn to positive results.

Possibilities

We begin this section by focusing on the *simple majority rule,* which is widely applied in many committees. According to this rule, alternative x, say, is majority-wise weakly preferred to another alternative, y, if the number of voters who find x at least as good as y is larger or equal to the number of voters who find y at least as good as x. The simple majority rule is an attractive decision mechanism since it treats both voters and alternatives equally (there is discrimination neither among voters nor among the issues to be decided upon—the latter property is called "neutrality"). A simple majority decision is also responsive to changes in voters' rankings, much more than the absolute majority rule, for example.

Why is the method of majority decision not a counterexample to Arrow's famous negative result? The answer is that under an unrestricted domain of individual orderings, the majority rule may generate so-called Condorcet cycles. In their simplest form, in the case of three voters, they come about if one person prefers x to y and y to z (and also, of course, x to z, due to transitivity), the second voter prefers y to z and z to x, and the third person prefers z to x and x to y. Simple majority counting establishes a social preference for x over y, for y over z, and for

z over x. In other words, there is a preference cycle on the aggregate level so that an Arrow-type social welfare function does not exist. For three alternatives and three voters, the probability for cyclical social preferences to occur is about 5.5%, which is not very large, but this probability increases with the number of voters and the number of alternatives, so it cannot be considered as *une quantité négligeable*.

The Welsh economist and political scientist Duncan Black has shown that if the individual preference rankings are "single-peaked" over each triple of alternatives, the simple majority rule yields a Condorcet winner—that is, a candidate who is majority-wise at least as good as every other candidate. If the number of voters is odd, the method of majority decision generates a social ordering (so that the rule becomes an Arrow social welfare function). Single-peakedness as a qualitative property on preference profiles can be taken literally. Each and every voter has a most preferred alternative (in terms of ordinal preferences), and on either side of this most preferred object, the person's preference decreases. Black himself thought in terms of the political spectrum (left–right). Of course, the far left and the far right voters show declining preferences only on one side of their respective peaks.

Amartya Sen generalized Black's condition to what he called "value restriction." For each triple of alternatives, there exists one alternative such that all voters agree that it is never the worst (case of single-peakedness), that it is never the best (case of single-troughed preferences), or that it is never in the middle between the other two. Again, a Condorcet winner exists under the simple majority rule, and this rule once more becomes an Arrow social welfare function for an odd number of voters.

Single-peaked preferences have another remarkable property. Under this domain restriction, the method of simple majority decision turns out to be strategy-proof. This was shown by Hervé Moulin very generally—the position of the so-called median voter balances deviating interests to the left and to the right of this voter's peak. A strategic misrepresentation would be against the very interest of those who are involved in the decision making.

Consider a social choice rule that specifies for any two alternatives x and y that x is socially at least as good as y if and only if it is not the case that all voters find y at least as good as x and there is at least one individual who finds y strictly better than x. Sen called this rule the *Pareto-extension rule*. It fulfils Arrow's conditions of unrestricted domain, weak Pareto, the independence condition, and nondictatorship, but it does not satisfy the requirement of full rationality of the social relation. The Pareto-extension rule is not transitive with respect to the indifference part of the social preference relation. Therefore, this rule is not a counterexample to Arrow's theorem either; it "only" constitutes a social decision function. The latter is always able to provide a nonempty choice set but not necessarily a social ordering. The reader realizes that subtle differences can matter a lot. The Pareto-extension rule may be viewed as unsatisfactory for social decision making, since whenever there is at least one person who strictly opposes the strict preference of the rest of society for x over y, say, the social outcome between x and y will be an equivalence or indifference. Such a person has been called a "weak dictator." A consequence of this weak dictatorship is that the Pareto-extension rule is largely unresponsive to changes in the underlying preference profile of the members of society.

The so-called Borda rank-order method, which without any doubt is the best-known rule from the class of general scoring functions, attaches ranks to positions. If there are m alternatives on which a social decision has to be made and if all alternatives are ranked in a strictly descending way by all voters, Borda proposed to attach the rank $m - 1$ to the top-ranked element, $m - 2$ to the second element from the top, and so on, and, finally, 0 to the bottom-ranked alternative. The element(s) with the highest aggregate rank sum is (are) socially chosen. Clearly, this rule uses much more (positional) information than the majority rule, for example. The latter just registers whether x, say, is preferred to y, or vice versa. The Borda rule also considers the "distance" between options—in other words, the number of positions of the alternatives between x and y.

Peyton Young showed that the Borda rule can be uniquely characterized by being neutral, consistent, and faithful, and having the cancellation property. Neutrality has already been explained in connection with the majority rule. Consistency makes the following requirement.

Imagine that the preference profile of society is split up into two disjoint subprofiles representing two disjoint sets of voters. If the intersection of the set of elements picked by the choice rule from the first

subprofile and the set of elements chosen from the second subprofile is nonempty, consistency requires that all elements in this intersection are identical to those that would have been picked by the choice rule if there had only been one profile, namely, the union of the two subprofiles. A scoring function is faithful if "socially most preferred" and "individually most preferred" have the same meaning when society comprises just one person. Note that if a scoring function is consistent and faithful, it satisfies the weak Pareto principle. Finally, the cancellation property requires that given any set of alternatives, if for all pairs of alternatives from this set, the number of voters preferring *a* to *b*, say, equals the number of voters with the opposite preference, then all elements from this set are equally chosen.

Clearly, the Borda winner can be different from the Condorcet winner. It can be shown that a Condorcet winner is never bottom ranked according to the Borda rule, and a Condorcet loser—a candidate who loses in pairwise contest against all other options—is never top ranked according to the Borda count. The equidistance between two adjacent scores is typical of the Borda rule but by no means necessary for general scoring functions. Nonlinear transformations of the Borda rank numbers, for example, can be introduced if there are plausible reasons for doing so.

Beyond Ordinal Noncomparability

Up to this very point, we have only been considering ordinal rankings both on the individual and the societal level, with no trace of interpersonal comparability. John Rawls's approach to justice, however, presupposes that we can compare levels of well-being (strictly speaking, in terms of so-called primary goods) across persons, so that we can, for example, say that person *i* is better off under state *x*, say, than person *j* is under *y*. This kind of interpersonal comparability is called *ordinal-level comparability*. Utilitarianism, which was made well known by Jeremy Bentham, a legal and social theorist, postulates utility of action according to the maxim of the greatest happiness of the greatest number possible. In more prosaic terms, utilitarianism either maximizes the aggregate sum of utilities over all persons concerned or the average utility per society. In either case, differences in utilities must be a meaningful concept, and these have to be compared across persons. The underlying concept here is the cardinal concept comparable to temperature or weight, where we also usually measure differences.

Apart from the informational requirement, namely, ordinal-level comparability in the Rawlsian approach versus comparability of utility differences in utilitarianism, Rawls focuses in his second principle of justice on the worst-off in society, whereas utilitarianism, as just stated, focuses on the aggregate sum of utilities, which is to be made as large as possible. Axiomatically speaking, both models of distributive justice have a lot in common. Both maxims satisfy the Pareto principle, independence of irrelevant alternatives, and anonymity. The point of bifurcation between both rules is a so-called equity axiom in the Rawlsian setup, with its single focus on the worst-off (strictly speaking in terms of an index of primary goods, such as income, wealth, opportunities, and self-respect).

The modern-day version of utilitarianism was formulated by John Harsanyi. His model, based on the Bayesian rationality postulates, employs the von Neumann–Morgenstern expected utility hypothesis. In one of his models, an ethical observer evaluates different policies for a particular society in an impartial manner, thereby determining the utilities that accrue to each member of society. This approach is very reminiscent of the role that an impartial spectator plays in Adam Smith's *Theory of Moral Sentiment*. Again, as in the case of de Borda and Condorcet, a fundamental idea bridged over two centuries, receiving a modern scaffolding, so to speak.

Finally, we would like to mention Sen's capability approach. For Sen, what defines freedom, autonomy, and well-being are the functionings of a person—her achievements—and not just the accumulation of primary goods as in Rawls's theory. What a person manages to do or be (e.g., being well nourished and well clothed, taking part in community life, having access to medical care) are functionings that are important for a person's life. The total number of functionings that are available to a person or household define the advantages of that person—her real opportunities. These make up the person's capability set.

Wulf Gaertner

See also Allais Paradox; Bargaining Theory; Capabilities; Collective Rationality; Decision Theory; Judgment

Aggregation and the Recursive Dilemma; Pareto Optimality; Preference; Rational Choice and Political Science; Rational Expectations; Sen's Paretian Liberal

Further Readings

Arrow, K. J. (1963). *Social choice and individual values* (2nd ed.). New York, NY: Wiley. (Original work published 1951)

Black, D. (1948). On the rationale of group decision making. *Journal of Political Economy, 56,* 23–34.

Gaertner, W. (2009). *A primer in social choice theory* (Rev. ed.). Oxford, England: Oxford University Press.

Gaertner, W., Pattanaik, P. K., & Suzumura, K. (1992). Individual rights revisited. *Economica, 59,* 161–177.

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica, 41,* 587–602.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparability of utility. *Journal of Political Economy, 63,* 309–321.

Moulin, H. (1980). On strategy-proofness and single-peakedness. *Public Choice, 35,* 437–455.

Rawls, J. (1971). *A theory of justice.* Cambridge, MA: Harvard University Press.

Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory, 10,* 187–217.

Sen, A. K. (1970). *Collective choice and social welfare.* San Francisco, CA: Holden-Day.

Sen, A. K. (1985). *Commodities and capabilities.* Amsterdam, Netherlands: North-Holland.

Smith, A. (1759). *The theory of moral sentiment.* London, England: Millar in the Strand.

Young, H. P. (1974). An axiomatization of Borda's rule. *Journal of Economic Theory, 9,* 43–52.

Social Cognition

The term *social cognition* has two primary referents. (1) As a subdomain within the parent discipline of social psychology, its mission is to apply the methods and models of cognitive psychology to study the mental representations underlying social behavior. (2) As a line of inquiry, heavily influenced by philosophical, neurological, and evolutionary (as well as) cognitive perspectives, its goal is to discover whether there are mental structures and processes unique to a person's knowledge of and interaction with social entities (as opposed to nonsocial objects, facts, etc.).

This entry examines these two areas of application of the notion of social cognition and ends by using memory—and in particular episodic memory, central to social interaction—as an example of certain methodological and ontological issues raised in social-scientific research. The discussion also highlights the importance of evolutionary perspectives in social cognition.

Social Cognition as a Subdiscipline of Social Psychology

Social cognition originally emerged as a subdivision within social psychology. Its defining features were that it combined (a) the parent discipline's concern with human social interaction with (b) an interest in applying models and techniques derived from the study of nonsocial cognition (e.g., memory for words) to explore the mental representations underlying social behavior. To appreciate the revolutionary nature of this convergence, it helps to review the context in which the interaction between social and cognitive psychology developed.

Starting in the late 1950s, hard-line positivism's informal ban on what the behaviorist movement had labeled "black box" psychology gradually gave way to the appreciation of the importance of inferring mental structures from observed behavior. This refocusing, first adopted by cognitive psychology in the early 1960s, did not enter social psychology until roughly a decade later. Emerging in the mid-1970s as a subdivision within social psychology, social cognition continued to focus on topics of interest to its parent domain, with the additional constraint that understanding social behavior must be grounded in an explicit appreciation of what the mind actually can and cannot do. Accordingly, social cognition proposed and tested hypotheses of mental functions (e.g., perception, memory, automaticity) presumed to underlie the overt behaviors of interest to social psychologists (e.g., impression formation, attitudes, stereotyping, conformity). This was achieved largely by borrowing measures (e.g., response latencies), procedures (e.g., priming and recall paradigms), and models (e.g., associative and computational) from cognitive psychology to serve as bridges with which to infer the underlying cognitive mechanisms responsible for interpersonal behavior. In placing a premium on inferring the content of the "black

box," social cognition helped position social psychology within mainstream cognitive science.

By the time a second *Handbook of Social Cognition* was issued in 1994, the parent field of social psychology was undergoing a gradual transition toward the "mental" side of human social behavior—thus adopting a shift toward theoretical commitments that had taken center stage in cognitive psychology 20 years earlier. Over the ensuing decades, there has been continual updating of cognitive representational theories (e.g., parallel distributed processing [PDP] models) and methods (e.g., functional magnetic resonance imaging). As the cognitive science approach became the standard paradigm across psychological domains, social cognition gradually morphed from an independent subdomain within social psychology to a set of methods, procedures, and theoretical orientations absorbed into the parent domain. Thus, the success of social cognition in aligning the study of social interaction with other areas in the psychological sciences rendered the need for a separate subfield within social psychology superfluous. Social cognition currently exists more as a set of procedures and assumptions within social psychology rather than as a separate subdivision.

Social Cognition as a Search for Cognitive Processes Unique to Social Behavior

For most of the past century, psychologists have been exploring the capabilities of mental systems, which are staggering. These mental systems can process a vast array of information, including much that is adaptively arbitrary and evolutionarily novel—from nonsense syllables and batting averages to chess moves and equations. But agnostically cataloging arbitrary samples out of the inexhaustible set of everything a memory system is capable of doing is not likely to lead to knowledge of its function. From an evolutionary perspective, the brain is a system designed to solve problems—that is, a system whose parts exist in their present form *because* that arrangement solves a problem recurrent in our ancestral environment.

A second referent of the term *social cognition* thus trades on the question "Have we evolved specialized neural structures and mental processes that enhance our ability to successfully navigate the complexities of our social world?" Current research suggests that the answer is "yes."

Take, for example, the human face. The face seems to be especially relevant to a host of basic social-cognitive process. It is the point of contact in the infant's very earliest social interactions; the smiles exchanged between an infant and its caregiver is the beginning of lifelong social bonds. Perceiving, identifying, and comprehending faces is absolutely basic to social interaction. We have to know who we are dealing with, what they are like, and how we relate to them, before we can interpret their behavior or plan our own. The face (even a stranger's) provides cues to the emotional state of the other person, as well as hints of other things, like deception, that are important in negotiating an interaction. Consistent with these considerations, several decades of study reveal that there are neural structures specifically devoted to face processing—structures that when impaired produce specific deficits in face perception (e.g., prosopagnosia, or "face blindness") but not perception more generally.

Theory of mind (TOM) is another cognitive process specifically devoted to social behavior. TOM enables its possessor to infer unobservable mental states (e.g., beliefs, goals, thoughts, motives) in others, thereby providing its owner a significant adaptive advantage in predicting and interpreting the behavior of others. This ability has no known nonsocial analog (save for the widespread anthropomorphism evidenced by our hominid ancestors) and evidences highly specific patterns of loss in neurological disorders characterized by patients' failure to participate normally in social communication and interaction (e.g., autism).

The list could be greatly expanded (e.g., mirror neurons, which respond both when an animal acts and when the animal observes the same action performed by another—accordingly, the neuron "mirrors" the behavior of the other).

Let us now consider an example of how evolutionary and philosophical considerations led researchers to rethink the origins of a classic form of memory whose status as a social adaptation became evident only after researchers began to ask questions about its evolutionary functions (rather than questions more traditionally posed by cognitive science that pertain to a system's capabilities).

Episodic memory is a subsystem of long-term memory held to consist of knowledge of a previously experienced event, along with an awareness that the event occurred in one's past. Episodic memory has traditionally been viewed as having evolved

in response to selection pressures posed by general environmental contingencies that the organism encountered—both social and nonsocial. However, evolutionary and philosophical considerations, confirmed by empirical testing, now make it appear that episodic memory is an adaptation specific to the social demands of the environment.

The fact that scientists have an abstract category—*episodic memory*—does not guarantee that an ontological correlate exists in the brain. If episodic memory is a genuine system in the biological sense, then it exists in its present form because that arrangement solved certain recurrent problems faced by the organism in its evolutionary past. Evolution does not produce by chance new phenotypic systems that are complex and functionally organized; rather, systems acquire their functional organization because they contributed to the organism's ability to survive and reproduce. In the case of episodic memory, these include, but are not limited to, keeping track of cooperative relationships (e.g., social exchange, cheater detection, coalitional allies) and reevaluating social knowledge in light of new evidence. Thus, episodic memory enables its owner to navigate more successfully in the complex world of human social interaction.

This is not to say that episodic memory is capable of performing only socially relevant tasks. Every system, by virtue of having a particular causal structure, is capable of doing an endless series of things that it was not designed to do. What research suggests is that episodic memory is the functional product of an evolved adaptation designed to facilitate interaction with other people.

In summary, a multitude of cognitive adaptations appear specific to social behavior. Social complexity being a hallmark of our species, it would be surprising if this were not the case. Adopting an evolutionary perspective as a heuristic for discovering such mechanisms is a promising direction for future social-cognitive research.

Stanley B. Klein

See also Behaviorism in Psychological Explanation; Cognitive Sciences; Evolutionary Psychology; Grounded Cognition and Social Interaction; Joint Attention and Social Cognition; Simulation Theory; Social Neuroscience; Social Perception; Theory Theory

Further Readings

Brewer, M. B., & Hewstone, M. (Eds.). (2004). *Social cognition.* Oxford, England: Blackwell.

Hamilton, D. L. (Ed.). (2005). *Social cognition: Key readings.* New York, NY: Psychology Press.

Fiske, S. T., & Taylor, S. E. (2007). *Social cognition: From brains to culture.* Hightstown, NJ: McGraw-Hill Higher Education.

Klein, S. B., Cosmides, L., Tooby, J., & Chance, S. (2002). Decisions and the evolution of memory: Multiple systems, multiple functions. *Psychological Review, 109,* 306–329.

Wyer, R. S., & Srull, T. K. (Eds.). (1994). *Handbook of social cognition* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Social Construction of Reality

This entry presents and discusses the once influential theory of Peter Berger (1929–) and Thomas Luckmann (1927–) on the social construction of reality, which appeared in their classic book by that title. The entry proceeds to review some critical points about this subject and points to recent developments.

Berger and Luckmann's Theory

When it first appeared in 1966, *The Social Construction of Reality* by Peter Berger and Thomas Luckmann made a powerful impact on sociology, for several reasons: It was, as its subtitle said, "A Treatise in the Sociology of Knowledge," an area largely neglected in the contemporary sociology of the day; it brought the name and work of the social scientist, philosopher, and lawyer Alfred Schütz (1899–1959; who had taught both authors at the New School for Social Research) to the awareness of many sociologists for the first time; and it was qualitative and humanistic in its perspective, in contrast to the structural-functional and quantitative sociology of the day. It thus provided considerable support for those advocating qualitative instead of quantitative methods in sociology and, by introducing Schütz and many of his concepts, provided an impetus for those interested in phenomenological approaches as well.

The authors had undertaken to examine knowledge, commonsense knowledge as held by the "man in the street," the ordinary member of society. Their mission was to carry out what Schütz had advocated for many years—that is, a serious, intense, and detailed study of whatever passed for knowledge among ordinary persons. Schütz referred to the work by Max Scheler, Karl Marx, and Karl Mannheim as the so-called sociology of knowledge.

By focusing the sociology of knowledge on everything that passes for knowledge in society, they would thereby achieve what Peter Berger called a "democratization" of the sociology of knowledge, rather than dealing with the history of ideas, or ideology, as examined by intellectuals.

Berger and Luckmann state clearly on the first page of their work that, for them, "knowledge" and "reality," when examined sociologically, are found to pertain to specific social contexts. Any analysis of the context—that is, the study of various aspects of a society—would have to take this into account. The taken-for-granted in one society (or even subuniverse) may not be the same in another.

They further advocated that any examination of empirical varieties of knowledge has also to consider how *any* body of knowledge comes to be socially established as "reality." Those who do research, therefore, would not judge the validity or invalidity of what is taken to be known (as reality) but would rather seek to discover and describe knowledge for a particular society and/or its various subdivisions—in other words, to analyze how knowledge is socially constructed.

By *social construction,* they meant the interactive processes engaged in by the man in the street together with others, which led to commonsense knowledge—that which is known and then taken for granted by them.

They themselves "bracket" epistemological and methodological questions concerning the validity of sociological analysis. They wish to focus on what is known as "real" by members of society in their ordinary, everyday lives and not in their own theorizing about it—not only the content of what is "known" by members of society but how the known, as objective, can also be *subjective.* In this undertaking, Berger and Luckmann borrow from or modify several influential sources, including Émile Durkheim, Karl Marx, Max Weber, and George Herbert Mead, to delineate what they consider the most relevant aspects of the sociology of knowledge. These are how reality is *interpreted,* what *subjective meanings* members hold as they develop the coherence of their world, and how subjectivations (of meanings) are objectivated to produce an intersubjective commonsense world. Persons are aware of multiple realities and that, for example, the world of dreams is different from the world of everyday life (what Schütz called the "paramount reality")—in other words, that there are different "realities" and that everyday reality appears as already "objectified" to them and to others.

The mode of interaction with others, the typificatory schemes that may be used, the language available for use in interaction are all sources of objectivation of experiences and may lead to an interpretation of these as "already there," formed independent of us. Nevertheless, we also see how reinterpretations and modifications, and also actions based on these, can be made by ourselves and others. They argue that the sum total of these typifications and of recurring interaction patterns results in a social structure that forms part of the reality of everyday life as we perceive it.

One may use typifications in interactions with others, as may they, and despite the increasing anonymity of others, one may also "know" how to act in taking them into account. Such anonymity may depend on the input or presence of interest and intimacy, both of which may vary over time, since one "knows" one's predecessors and contemporaries better than one's successors (or even consociates).

Berger and Luckmann go on to discuss habitualization and its role in the formation of institutions, institutionalization and its being experienced as an objective reality, and the human or social production of a world of meanings shared with others as social relationships are formed.

According to Berger and Luckmann, sociological investigations of a phenomenon may examine how its meanings were socially produced—that is, how they came to be produced by and for the ordinary man in the street. The researcher may focus on the various aspects of production with the aim of revealing which types of actors, which settings, and which formulations or linguistic forms were/are used in their ongoing production.

One aim may be to offer critical perspectives on these descriptions and analyses in order to reveal the importance of existing persons (contemporaries) or previously present predecessors to the construction of meanings. Part of the aim may be to change the meanings and enhance the efforts of those attempting to modify or reinterpret the meanings of the phenomenon or who propose different linguistic usages for describing and analyzing the phenomenon.

The fact that Berger and Luckmann had presented a new way of examining knowledge, how they had utilized Schütz (together with Durkheim, Marx, Weber, and Mead) to do this and how they had focused on commonsense knowledge and said that the observer did not have to assess the validity of the claim to certainty with regard to the existence, shape, or meanings of phenomena, did not result in extensive studies of commonsense knowledge and the taken-for-granted or in multiple realities or in most of the conceptualizations they had introduced.

Berger and Luckmann did *not* claim that everything is socially constructed or that everything is a social construct—that is, they did not embrace a universal social constructionism.

They did say that institutional order is developed and maintained by members' knowledge of it, often by metaphor, maxims, morals, myths, and so on. Any analysis would reveal the recipe for knowledge that can provide rules of conduct for maintaining the institutional order. A focus on institutional order would show what rules of conduct were sanctioned, what was held as true, as belief, as well as what social controls were utilized for its maintenance/preservation. Deviance from these would thus be recognized, and efforts would be made to sanction those who did not conform.

A social world thus constituted would then be accepted as "the world" and utilized to assess any alternative actions or judgments. Such knowledge could then be internalized, and reference to external, objectivated structures could be assessed in relation to it.

Knowledge, in their view, provides a means for production of an objective world, uses language to order those objects that will be considered part of social reality, and, when internalized through socialization, can be accepted as valid or true. Knowledge thus serves not only to maintain but also to produce on an ongoing basis the reality accepted by members of that society.

Much of this knowledge is sedimented, remaining in memory and not requiring the individual to have to examine everything at hand. Language is the principal means of intersubjective sedimentation, which then enables all parties who have experienced particular phenomena as well as those who use the same linguistic meanings to share objectivations in the same collectivity.

They may become legitimated in the process of achieving a place in the stock of knowledge, such that one can say, "We know that there is this aspect to reality." The knowledge of reality becomes a shared and transmittable knowledge, and succeeding generations may even be seen as representative of such viewpoints. That is, a collectivity may be known for its traditions, the knowledge shared by its members, and the views and definitions of the "reality" that they share.

As Berger and Luckmann saw it, there is a dialectical relation between knowledge and the social sedimentations held by the members of a collectivity. Knowledge may be socially produced, but it also may serve to produce further social change in the meanings of the objectivated world.

Each subuniverse and institutional order may become objectified, that is, apprehended as though it is not a human product. A loss of awareness may lead persons to think that the reified social reality cannot be changed, as though one has no control over it. The objectivated meanings that lead to *reification* may be forgotten, and paradoxically, persons may interpret their own situations as being shaped by the objectivated, or reified, reality. In losing awareness of their own role in institutionalization, objectivation, and reification, they may believe that this is the way things were meant to be, that divine guidance has produced them, or that this is only natural. Here, we can see that any number of interpretations or explanations with regard to sources or causes may be made.

Furthermore, they emphasized that any theory must be *dialectical* in the sense that it should continue to examine the relation between institutionalization (and reification) and the actions and beliefs of human actors as these are continually involved in their maintenance and production. A loss of this perspective would mean that the theorist/observer would no longer see the *social* construction of reality—how "reality" is socially produced.

Critical Assessment and Current Developments

The book, *The Social Construction of Reality,* is indeed what its subtitle proposes, "A Treatise on the Sociology of Knowledge." However, adoption of the phrase, and supposedly the perspective that Berger and Luckmann presented, has led to many varied interpretations, some of them wrong or misguided, and to the identification of "social constructionism" as a perspective that is critical of existing ideas, beliefs, or institutions and, in general, conventional meanings in sociology or whatever is being examined or written about.

These developments and misconstruals of the meanings intended in the book have led Berger to "shy away" from constructionism and Luckmann to say that whenever he hears the terms *social construction* or *constructionism,* he "runs for cover." In fact, some versions of social construction(ism) do not even reference Berger and Luckmann's original formulation; for example, the *Stanford Encyclopedia of Philosophy* considers it a "metaphor."

In a close examination of the meaning and use of the terms *the social construction of reality* and *constructionism,* Ian Hacking notes that many studies have used them in their titles, ranging from the social construction of authorship, to brotherhood, danger, emotions, facts, and more, including the last on his list, Zulu nationalism. In Hacking's book on the "social construction of _____", *X* represents whatever it is that is being studied/written about, and he characterizes much of this work as having primary value for political activists and those involved in debates about race, gender, culture, or science.

Hacking distinguishes two different usages and referents (of whatever *X* is): a primary way "for raising consciousness" and a way of being critical of the status quo, that is, *X* need not have existed, is quite bad as it is, and we would be much better off if *X* were eliminated or changed. The main idea is that *X* is not inevitable and does not have a particular character or nature independent of human beings. Rather, it evolved from social situations, forces, and historical conditions that, if varied, could have resulted in a different character of *X*. This is not to argue for relativism but rather that doing social construction analysis will enable one to show how the taken-for-granted meaning of *X* emerged over time to its present state.

Social construction may then be about objects and/or ideas and the interaction between them. Objects are, in commonsense knowledge, taken for granted and assumed to be what they (currently) appear to be or are defined as being. These may be considered to be essential or inevitable for all human beings and at all times—that is, a universal constructionism.

Ideas and concepts may also be taken for granted and may appear as inevitable, whereas a closer examination may reveal how certain interpretations are used, how they have developed, and how they are related to social and historical conditions. Such revelations may be used to offer alternatives or to criticize existing interpretations and, clearly, to reveal how they are *socially* constructed.

Abstractions, or what Hacking calls "elevator words," are also shown to have similar usages and characteristics.

Social constructionism is examined by Hacking not as a theory concerning the sociology of knowledge but primarily as a critical perspective that seeks to reveal how existing ideas and meanings have evolved. Such revelations could then lead to social change—that is, to alternative interpretations (meanings).

Concluding Remarks

The idea of social construction has evolved and even changed since the original formulation by Peter Berger and Thomas Luckmann. An examination of their theory and its Schützian background reveals that they were involved in complex formulations of relevance to a theory of the sociology of knowledge. Distortions and misinterpretations have come in a variety of guises, but the original formulation remains more complex than those who claim to do social construction have ever recognized.

George Psathas

See also Ethnomethodology; Intersubjectivity; Metaphor; Philosophy of Sociology, History of; Reification; Social Constructivism; Sociology of Knowledge and Science; Weber's *Verstehende* Approach

Further Readings

Berger, P. (1992). Reflections on the twenty fifth anniversary of the social construction of reality. *Perspectives, 15*(2), 1–4.

Berger, P., & Luckmann, T. (1967). *The social construction of reality: A treatise in the sociology of knowledge.* Garden City, NY: Anchor Books.
Hacking, I. (1999). *The social construction of what?* Cambridge, MA: Harvard University Press.
Luckmann, T. (1992). Social construction and after. *Perspectives, 15*(2), 4–6.
Mannheim, K. (1956). *Ideology and utopia: An introduction to the sociology of knowledge.* New York, NY: Harcourt Brace & World.
Schütz, A. (1962). *Collected papers* (Vol. 1). The Hague, Netherlands: Martinus Nijhoff.
Schütz, A. (1964). *Collected papers* (Vol. 2). The Hague, Netherlands: Martinus Nijhoff.

Social Constructivism

Social constructivism with respect to a given phenomenon is the view that the latter does not possess an independent existence but is "constructed"—that is, generated and maintained through collective human action, thought, discourse, or other social practices. Constructivism is thus antithetical to realism with respect to the same entities, which would ascribe autonomous existence to them. Social constructivism is primarily a position in philosophy of science, but it also informs and inspires a number of recently influential schools within the empirical social sciences themselves.

This entry introduces the various meanings and uses of *social constructivism* and reviews four main positions with regard to the ontological and epistemic versions of social constructivism concerning the social world and the physical world, respectively.

The term *social construction* was first used by Peter Berger and Thomas Luckmann in their 1966 work, *The Social Construction of Reality.* In this book, what is declared to be a construction is, albeit ambiguously, the social realm. Subsequently, the constructivist stance has achieved a widespread following and has been extended to other spheres. While social constructivism as defined above is trivially true for many human phenomena and artifacts, it becomes controversial when applied to areas of reality, or features thereof, that are normally held to exist autonomously.

Social constructivist positions often harbor normative or ideological overtones. As stressed by Ian Hacking, claims that a certain phenomenon is "constructed" carry the implication that the latter is not part of "the natural order of things"; it is not eternal and immutable, let alone necessary. Hence, such claims are often raised in connection with efforts to effect societal changes. Favorite areas in which constructivist claims have been raised in order to promote normative agendas pertain to divisions of gender or race.

Underneath its general nominal definition, social constructivism is a very diverse intellectual trend, and its adherents across the fields would not often claim any communality. Moreover, there is often considerable disagreement between constructivists within any particular field, as they draw upon many and diverse philosophical traditions ranging from phenomenology and Marxism (both of which inspired Berger and Luckmann's seminal effort) to philosophy of language. Still, the label is useful in identifying a characteristic collectivist and anti-realist mode of thinking, local or global, that has recently enjoyed considerable popularity.

Two intersecting distinctions among constructivist claims will help create an overview of this complex field.

The first distinction is that between the *material or physical* sphere and the *social and human* sphere, as objects of construction. The second distinction is that between *ontological* and *epistemic* constructivism. According to ontological constructivism, the object of construction is the world itself, while according to epistemic constructivism, it is our knowledge of (beliefs about) the world, including our scientific knowledge. (Knowledge, of course, also belongs to the world but constitutes a special subpart that is subject to evaluative, rational constraints and, hence, may usefully be placed in a separate category.) The import of claiming that (scientific) knowledge is a social construction is that it is shaped by societal forces rather than through a process of "tracking" the reality that is its nominal object.

When the two distinctions are combined, social constructivist positions may be organized as shown in Table 1.

Ontological Constructivism With Respect to the Social and Human World

That the social world depends for its existence on the thoughts, language, and actions of its inhabitants is hardly controversial; what is of interest is rather

Table 1 Varieties of Social Constructivism

	Ontology	*Epistemology*
The material (physical) world	The material world is constructed by collective human thought and practice	Our knowledge, including our scientific knowledge, about the material world is constructed by collective human thought and practice
The social and human world	The social and human world is constructed by collective human thought and practice	Our knowledge, including our scientific knowledge, about the social and human world is constructed by collective human thought and practice

Source: Author.

what precise mechanisms are held responsible for its generation. In recent discussions, emphasis has been lain on the linguistic sources of such construction, a development that is in line with the general "linguistic turn" in 20th-century philosophy and social science. Two major traditions may be distinguished, which may be referred to as the analytical and continental arguments, respectively.

1. The *analytical argument* is represented by authors such as Margaret Gilbert, Raimo Tuomela, and John Searle, who hold society to be constructed through the collective intentionality ("we-intentionality") of its members. Searle's version in particular utilizes and extends insights from speech act theory. The most important social facts are institutional rather than "brute" (i.e., natural), and the analysis of such facts provided by speech act theory can be extended to all societal institutions. A core notion is that of constitutive rules, which state that some person, thing, or action *counts as* something else under certain conditions. They thereby acquire a certain status that enables them to serve a certain social function, since various normative powers (rights, duties, obligations) flow from the imputed status. These powers motivate specific collective actions and thus instigate and direct social interactions. When a body of constitutive rules defining status functions is recognized in a given population (a case of collective intentionality), collective institutions and, thus, social reality are created. Money is the classical example, in which a certain power—"purchasing power"—is bestowed upon pieces of paper that are in themselves valueless.

This type of constructivism has affinities with classical liberalist political philosophy and its conception of society as formed and legitimized through a social contract. In their later work, Searle and Toumela have attempted to link their constructivism with aspects of this tradition, such as the notions of political power and human rights.

2. According to the continental argument (so called), social reality is also generated by the linguistic categories in which we talk about it, but in a less specific manner. The philosophical underpinnings of this argument are typically derived from Michel Foucault's writings. The fundamental line of thought harks back to structuralism, which holds that language (*langue*) is an arbitrary structure of differences that, when manifested in concrete discourses (*parole*), generate a semantic system—that is, a system of categories under which objects in the world are subsumed. When these objects are human actions and their products, a social reality ensues, structured according to the categories of that particular discourse. In Foucault and other continental authors, these ideas are taken through a "poststructuralist turn," which emphasizes that the linguistic structures are not completely rigid but dynamic and subject to transformations through discourses articulating social interests—interests that, however, are themselves transformed in the process. In empirical social science, this view of social reality has, for example, been adopted in the influential theory of democratic politics by the political scientists Ernesto Laclau and Chantal Mouffe. Laclau and Mouffe, along with other representatives of the continental argument, draw upon Wittgensteinian ideas as well, in particular the notion of "language–games," with its implication of an irresolvable fusion between language and social practice that makes language constitutive of social reality.

There are representatives of the same mode of thought in the anglophone academic world as well. Within philosophy of social science, it has been

articulated by Peter Winch, while Ian Hacking adopts a similar position, albeit on a rather more concrete and empirical level. This position was anticipated in a less philosophical vein by the labeling theory in the sociology of the 1960s and 1970s, for example, in the work of Howard Becker.

Epistemic Social Constructivism With Respect to the Social World

Epistemic social constructivism is, in particular, the position that social-scientific knowledge is shaped by societal forces. The classical example is Marxism, which asserts that knowledge concerning the "superstructure" of society is shaped by the class interests of societal subjects; this is in contrast to natural science, which is held to be resistant to such influence. Marx's view paradigmatically manifests the normative aspect of constructivism in that the demonstration of the ideological nature of societal knowledge is meant to generate "critical consciousness" with respect to the latter. Another classical example of epistemic constructivism is Max Weber's claim that the concepts in which we describe the social world, in the form of "ideal types," are shaped by societal "values." In recent social science, a constructivist stance is represented by "discourse analysis," which holds that the semantic system in which societal knowledge is articulated (if not its detailed content) is shaped by social interests and serves to promote the latter. The representatives of this position often go on to draw ontological implications from this argument (cf. Laclau and Mouffe above).

Epistemic Constructivism With Respect to the Physical World

This stance has been the hallmark of the recently influential discipline of *science studies,* a radical outgrowth of classical sociology of science. The founding representatives of the discipline are David Bloor (a contributor to this encyclopedia) and Barry Barnes, who are core members of the so-called Edinburgh school. According to the school's Strong Program, as opposed to the weak program of traditional sociology of science (vide Robert Merton), theories even within the natural sciences are generated and shaped by social interests (a claim that extends the Marxist tradition) or by general societal but historically variable modes of thought (the Durkheimian tradition).

This position is in part based on a number of celebrated case studies, but it also draws support from familiar critical arguments within analytical philosophy of science. Chief among these is the instrumentalism of authors such as Willard Van Orman Quine and Baas van Fraassen. While these authors are satisfied with vague references to "pragmatic considerations" in specifying what forces decide the choice of theoretical models in the absence of a decisive influence from reality itself, science studies focus upon precisely this question and deliver the answer: societal forces. Another important influence is Wittgenstein's "rule-following argument" from the *Philosophical Investigations,* which is held to show that the correct application of all descriptive terms, including those of science, is a matter of social agreement and is, hence, subject to the social interests that shape such agreements.

Ontological Constructivism With Respect to the Physical World

Epistemic instrumentalism with respect to theories of natural science is often pushed to an extreme within the field of science studies. The reality toward which natural scientific knowledge is directed is claimed to exert no epistemic pressure whatsoever upon our scientific theories; hence, the idea that experimental testing can push us toward an ever more accurate depiction of physical reality is dismissed. A radical ontological constructivism ensues, which holds that theoretical entities such as atoms and black holes are purely pragmatic posits that only exist within the conceptual frameworks and societal practices in which they are embedded. Their ontological status is thus akin, for example, to that of money, which does indeed exist within a given monetary system, where slips of paper really do possess exchange value but have no existence outside the system. A position along these lines is adopted, for example, by Bruno Latour in his actor-network theory (although, in later works, Latour objects to being labeled a *social* constructivist, since the agents of construction include nonhuman "actants," e.g., scientific measuring apparatus).

Within science studies, too, the constructivist stance often goes along with a normative agenda that is normally tacit but sometimes quite explicit (e.g., in Latour's work). The classical conception of natural science with its realist ontology and rationalist epistemology is held to bestow an illegitimate societal

privilege upon scientists; by contrast, a social constructivist interpretation is thought to lead to a more democratic way for science to function in society.

Finn Collin

See also Actor-Network Theory; Collective Intentionality; Instrumentalism of Scientific Theories and Constructive Empiricism; Language-Games and Forms of Life; Realism and Anti-Realism in the Social Sciences; Searle and the Construction of Social Reality; Social Construction of Reality; Social Ontology, Recent Theories of; Social Studies of Science and Technology; Strong Program in the Sociology of Scientific Knowledge; Structuralism and Poststructuralism

Further Readings

Barnes, B., Bloor, D., & Henry, J. (1996). *Scientific knowledge.* London, England: Athlone.

Berger, P., & Luckmann, T. (1966). *The social construction of reality.* Garden City, NY: Anchor Books.

Collin, F. (1993). Social constructivism without paradox. *Danish Yearbook of Philosophy, 28,* 24–46.

Collin, F. (2011). *Science studies as naturalized philosophy.* Dordrecht, Netherlands: Springer.

Hacking, I. (1999). *The social construction of what?* Cambridge, MA: Harvard University Press.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society.* Cambridge, MA: Harvard University Press.

Searle, J. (1995). *The construction of social reality.* New York, NY: Free Press.

Tuomela, R. (2007). *The philosophy of sociality.* Oxford, England: Oxford University Press.

Social Contract Theories

This entry looks into the philosophical issues, especially regarding rationality, behind contractarian theories prominent in political and moral theory as well as in certain social sciences, principally in economics. It does not purport to give a historical account of all social contract theories.

Origins

A social contract account of normative rules or principles is one that seeks to justify moral or political institutions by reference to the idea of some sort of rational agreement between individuals. Such an account was first introduced in a famous passage from Plato's dialogue:

> They say that to do wrong is naturally good, to be wronged is bad, but the suffering of injury so far exceeds in badness the good of inflicting it that when men have done wrong to each other and suffered it, and have a taste for both, those who are unable to avoid the latter and practice the former decide that it is profitable to come to an agreement with each other neither to inflict injury nor to suffer it. As a result they begin to make laws and covenants, and the law's commands they call lawful and just. This, they say, is the origin and essence of justice; it stands between the best and the worst, the best thing to do wrong without paying the penalty and the worse to be wronged without the power of revenge. The just is the mean between the two. (*Republic,* 1974, 358e2–359b1)

This is, of course, a very cynical description of the theory in question. A noncynical but all too brief account is to be found in Epicurus, the Hellenistic philosopher.

The contract is only a second-best solution—not the one that anyone would prefer if they had the power to act unjustly toward others with impunity. Moreover, it implicitly suggests that the agreement reached need not be egalitarian in any meaningful sense. Some may be more powerful than others and so are in a position to threaten others who are not so powerful, unless the latter obtain special privileges in the agreement reached. That is, the contract that results may favor some more than others. Moreover, there is a natural disconnect between what a person might agree to in a general way and how he or she will choose to act in various specific circumstances. This becomes clear in the *Republic* when one of the protagonists of the dialogue, Glaucon, goes on to tell the story of the shepherd Gyges, who finds a ring that makes him invisible and proceeds to use the power it gives him to triumph unjustly over others.

In this way, Glaucon neatly argues, in addition, for the contemporary view, espoused by many economists, that such contracts are unavoidably subject to what is known as a *free rider* problem: Persons may be "willing" to agree, in principle, to certain restrictions, but in the absence of effective surveillance and enforcement mechanisms, they will be disposed to violate the rules when they can get away

with it. They will "free ride"—that is, secure the benefits of the agreement without paying its costs.

Modern Times and Critical Issues

What would a "noncynical" account of contractarianism look like? In the modern period, the clearest version of contractarianism is to be found in the writings of Thomas Hobbes (17th century), who argues in his famous work *Leviathan* (1651) that it is rational for persons to accept an extensive set of "laws of nature" that specify how they are to interact with one another. Even Hobbes, however, thinks, as does Glaucon, that such an arrangement must be backed up by an effective enforcement mechanism—the Leviathan.

Care must be taken here, however. In the famous "fools" passage of *Leviathan* (I, 5), Hobbes insists that where the other party has already performed, it would be against reason for one not to perform, even in the case where there is no Leviathan.

There are a number of issues that must be addressed if the contractarian view is to have any plausibility. In the first place, to justify a particular moral or political institution, one must presumably show that it is characterized by a rule or principle to which it would be rational for persons to agree to conform their behavior. However, is the reference here to an actual or a merely hypothetical agreement between persons?

It is not clear whether it is to an actual agreement. There is little talk among contractarians about certain rules or principles being binding because persons have *in fact* "contracted" to abide by them. However, if it is a hypothetical agreement—what persons would agree to under some specified set of conditions—one faces the objection raised in our times by the legal philosopher Ronald Dworkin (1977) that a hypothetical contract is "not simply a pale form of an actual contract; it is no contract at all" (p. 151). Moreover, just what are the relevant conditions? And what ensures that an agreement on some rule or principle under the specified conditions is also binding under any *other* set of conditions?

Another issue concerns the significance of the qualifier *rational*. The only clear and uncontroversial sense of *rational* here would seem to be "*instrumentally* rational": the doing of that which is the necessary means to the realization of one's objectives. But it is not clear that this sense of *rational* is relevant. We are presumably talking here about what it is rational for a group of interacting persons to do, where there is no assumption that they all have the same objective.

Another problem is that it might be "rational" to agree to a certain arrangement between oneself and others, but it *also* might be rational to avoid doing one's part as much as possible. This is the issue to which Glaucon refers in his discussion of the Ring of Gyges, and as we have seen, it amounts to what in the contemporary literature of economics is known as the "free rider" problem—that is, whether a given participant ought to abide by a rule or principle when others have or can be expected to do their part.

At a deeper level yet, on a social contract theory, reference to a "contract" or an "agreement" is perhaps best treated as really having only a *metaphorical* meaning, there being nothing in the theory that requires there to be either an actual or a hypothetical agreement. The social contract theory, on this account, is simply the view that it would be (instrumentally?) rational for persons to put into practice a particular kind of normative arrangement between persons—one that specifies some set of rights and duties for each of the agents. Moreover, by "put into practice," one is to understand that each of the persons in question should be committed to act in accordance with its (normative) dictates. On this interpretation of the theory, the arrangement would not be subject to the free rider problem, at least to the extent that the participants in the arrangement are instrumentally rational.

The problem of how to interpret "instrumental" in this context might be worked out along the following lines. We need to isolate a special sense of what best promotes the objectives of the agents involved, namely, one that applies specifically to interactive situations in which the persons involved do not have the same set of similarly prioritized objectives but are committed to disparate objectives. In such a situation, it could be argued, a failure on the part of the individuals involved to carefully coordinate their activities in some specific way could easily lead to a situation in which everyone does less well, in terms of the furthering by each of his or her objectives, than they could do if they were to coordinate their actions. On this view, if the manner in which the several agents interact leaves each worse off than they could have been, then the arrangement

in question is *not* instrumentally rational. We may suppose, then, that in such an interactive situation, the unanimous failure to realize as much as each could have realized counts as a failure to achieve an instrumentally rational outcome. Alternatively put, the question of the instrumental rationality of the arrangement arises whenever it can be shown that the coordination is not *maximally* mutually advantageous—that is, when there is some alternative that would benefit each even more.

This account of instrumentally rational interaction would also need to involve an important qualifier. A person would not be bound to act in accordance with a contract's normative requirements unless the person has assurance that others (at least most others) are also committed to act in accordance with those same requirements. For most contractarians, including Hobbes, it would be fully consistent with such an arrangement that some persons—specifically many of those who are less than fully rational—are committed to act in accordance with its requirements *only* because the arrangement provides motivating penalties for those who would be—by virtue of their not being fully rational—disposed to otherwise "free ride" on the efforts of others. Such penalties, of course, would not be perfectly motivating: They would work only for those who are not truly irrational. They would impose costs that a person would seek to (and could) avoid, if he or she were at least somewhat rational, but that will not move a thoroughly irrational person.

What about the content of what could be shown to be (in the sense just explained) an instrumentally rational arrangement between persons? Is the distinction between *moral* and *political* rights and duties useful here? This is doubtful. There is no reason to suppose that what is at issue here pertains to moral rather than political, or political rather than moral, arrangements. Indeed, the distinction is not very easy to make out in this context. Again, there may be many "moral" rules, and also many "political" rules, that cannot be shown to be maximally mutually advantageous (in the sense introduced above)—many rules, that is, that it may well be instrumentally rational for some but not necessarily for all persons to adopt. Arrangements that prohibit cruelty toward animals would seem to be a good example of this. One may have to appeal to special additional assumptions to establish that it is instrumentally rational to prohibit cruelty to animals. It must not be supposed, then, that a "contractarian" account can be given of everything that persons regard as normatively binding.

Where such an approach comes into its own, however, is with regard to rules or principles governing rights and duties that relate to what in our times the political philosopher John Rawls, for example, described as *primary goods*. These are goods such as income and wealth, as well as protections from various forms of interference by others and protections from a wide range of the vicissitudes that arise (especially natural catastrophes and diseases) from various events over which persons typically have very imperfect control. They can be characterized as primary goods by virtue of their being goods that virtually any person will value, regardless of the specific final objectives that they pursue. The place of these primary goods is perhaps most interestingly illuminated by a remark of Hobbes (1996), who argues in the *Leviathan* that

> felicity is a continual progress of the desire from one object to another, the attaining of the former being still but the way to the latter. The cause whereof is that the object of man's desire is not to enjoy once only and for one instant of time, but to assure forever the way of his future desire. And therefore the voluntary actions and inclinations of all men tend, not only to the procuring, but also the assuring of a contented life. (Pt. I, chap. 11, pp. 47–48)

It can be argued that possessing Rawls's primary goods makes possible not merely procuring but ensuring what is needed for a contented life.

John Rawls and the economist John C. Harsanyi are often taken to represent contemporary examples of the contractarian position. However, both are preoccupied with trying to show that under a very special set of circumstances—namely, conditions of radical uncertainty—it would be rational to adopt this or that principle of justice, and one is bound to wonder if this conclusion extends to conditions where such uncertainty does not obtain. Contemporary versions of contractarianism that are not subject to this objection are to be found in James Buchanan's *The Limits of Liberty* and also in David Gauthier's *Morals by Agreement*. Indeed, among contemporary philosophers, Gauthier offers perhaps both the most comprehensive and promising way to think about a contractarian justification of certain moral or political rules and principles.

To be sure, Gauthier's account is subject to another objection to contractarianism, to which I have already alluded: He supposes that the basic contract would be shaped by the relative bargaining power of the parties involved. Rawls's and Harsanyi's versions avoid this problem, for they insist that one must consider what persons would do when a decision must be reached under conditions of radical uncertainty (where one's relative power vis-à-vis others is unknown). A more convincing approach, perhaps, would be to argue, as this author does, that there would be significant negative long-range consequences for all if persons were to continually press whatever relative advantages they have over others.

Edward McClennen

See also Bargaining Theory; Collective Rationality; Common Goods; Conventions, Logic of; Cooperation/Coordination; Cost–Benefit Analysis; Law, Social Phenomenon of; Normativity; Promises and Agreements; Public Goods; Rational Expectations; Rationality and Social Explanation; Reflective Equilibrium; Social Choice Theory; Social Conventions; Social Norms; Social Rules

Further Readings

Buchanan, J. (1975). *The limits of liberty.* Chicago, IL: University of Chicago Press.

Dowrkin, R. (1977). Justice and rights. In *Taking rights seriously* (pp. 150–183). Cambridge, MA: Harvard University Press.

Epicurus. (1964). *Epicurus: Letters principal doctrines and Vatican sayings* (R. M. Geer, Trans.). Englewood Cliffs, NJ: Prentice Hall.

Gauthier, D. (1986). *Morals by agreement.* Oxford, England: Clarendon Press.

Harsanyi, J. C. (1953). Cardinal utility in welfare economics and the theory of risk taking. *Journal of Political Economy, 61,* 434–435.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy, 63,* 309–321.

Hobbes, T. (1996). *Leviathan or the matter: Forme & power of a common-wealth ecclesiasticall and Civill* (R. Tuck, Ed. & Trans.). Cambridge, England: Cambridge University Press.

McClennen, E. (2012). Rational cooperation. *Synthese, 187*(1), 65–93.

Morris, C. W. (1996). A contractarian account of moral justification. In W. Sinnott-Armstrong & M. Timmons (Eds.), *Moral knowledge? New readings in moral epistemology* (pp. 215–242). Oxford, England: Oxford University Press.

Plato. (1974). *The republic* (G. M. A. Grube, Trans.). Indianapolis, IN: Hackett.

Rawls, J. (1958). Justice as fairness. *Philosophical Review, 57,* 164–194.

Rawls, J. (1971). *A theory of justice.* Cambridge, MA: Harvard University Press.

Social Conventions

This entry reviews the major theoretical accounts of social conventions. It first introduces briefly their philosophical origin and then goes on to present the first contemporary systematic account of social conventions in terms of game theory. This is followed by an overview of some of the main theories and uses of social conventions in current work in the philosophy of social science.

Introduction

Many of our everyday social interactions are regulated by *conventions*. Eating manners, the kind of clothes we wear at the office, and the side of the road on which we drive are a few mundane examples. Roughly, a social convention is *a customary, arbitrary, and self-enforcing* rule of behavior that is generally followed and expected to be followed in a group or in a society at large. When a social convention is established, everybody behaves in a quasi-agreed-upon way, even if they did not in fact explicitly agree to do so. A social convention can thus be seen as a kind of *tacit* agreement that has evolved out of a history of previous interactions.

The study of social convention is relevant for the social sciences since much of social order can in fact be explained in terms of conventions and, thus, as social regularities that emerge and are sustained without the need of centralized planning and external enforcement by the state. In philosophy, the notion of social convention is appealing especially to those who aim to formulate naturalistic theories of normative phenomena in general (i.e., obligation, law) and of morality in particular.

Origin: David Hume

The 18th-century Scottish philosopher David Hume was the first to point to the relevance of conventional regularities to addressing both of these issues. Aiming in particular to demystify the nature of property and justice, Hume suggested that a convention corresponds to a pattern of mutually beneficial behavior that a group of agents follow when they know that such a pattern is mutually beneficial and that they expect each other to follow this pattern instead of another.

First Systematic Account: David Lewis

In modern times, the Humean approach to conventions has been revived by the philosopher David Lewis, whose theory clarifies the customary, arbitrary, and self-enforcing nature of conventions. Adopting a *game-theoretic* approach, Lewis proposed that a convention is a solution to a *coordination problem* arising in recurrent interactions. A coordination problem is considered as a situation characterized by at least two coordination equilibria. A coordination equilibrium is a combination of actions—one for each player—in which each player is strictly motivated to perform his component of the combination, conditional on his believing that the other players will perform theirs. Moreover, there exists at least one alternative combination of actions that has the same property. Finally, for each player, if a player performs his share of the combination, he prefers that the other players perform theirs. When an interaction contains at least two coordination equilibria and when coincidence of interests between the players prevails, the players are facing a coordination problem.

A classical example of a coordination problem is that of choosing the same side of the road in order to drive safely. If, in a society, a regularity in behavior in which each individual picks his share of a coordination equilibrium is established, then, according to Lewis's definition, this regularity is a convention. Since a conventional regularity is sustained if there is a system of concordant mutual expectations of conformity, a crucial component of any theory of convention is explaining the origins of these concordant mutual expectations. According to Lewis, the source of these mutual expectations is *precedent:* If the agents have a shared acquaintance with instances of successful coordination in a class of similar situations in the past, they will project this pattern into the future. Precedent is seen as a source of one kind of salience, which makes one coordination equilibrium a focal point and thus prominent with respect to any possible alternative.

Contemporary Approaches

The focus of Lewis's theory is mainly on how conventions, once established, reproduce themselves. A compatible but complementary approach addresses the problem of how conventional regularities emerge in the first place. Combining insights coming from theoretical biology, Robert Sugden, for instance, has employed *evolutionary game theory* to study the origins of conventions. The most general mechanism that has been suggested to explain their evolution is that of *symmetry breaking*. Avoiding collisions at a crossroad, for instance, requires a rule that specifies who is supposed to stop and who is supposed to move forward. Since, however, in the absence of any convention the positions of all the drivers are symmetrical, by observing each other's behavior, the evolutionary dynamics cannot converge on one of the coordination equilibria. However, if the players can also discriminate among contextual features of their situation, they might exploit an arbitrary symmetry to solve this problem. For instance, if drivers condition their behavior on who is coming from the right, they might evolve a convention that assigns priority to those coming from that side of the road. Thus, arbitrary cues can boost the evolution of arbitrary regularities.

Though the importance of conventions in solving coordination problems has been exploited in many areas of the social sciences (from economics to linguistics and law), limiting the role of conventions only to situations in which the interests of the players coincide is indeed an undue restriction. Actually, theoretical models in biology, economics, and philosophy have shown that together with conventions of coordination, conventions of *partial conflict* can also emerge and stabilize. Robert Sugden and Brian Skyrms have shown, for instance, that a convention of partial conflict in which property rights are assigned to the first person to take possession of a previously unowned item can evolve by exploiting the same symmetry-breaking mechanism sketched above.

Finally, even if a social convention is often regarded as a mere regularity in behavior, it has often been suggested that it also has a *normative* force. Margaret Gilbert, for instance, has argued that a social convention has an intrinsic normativity that can be accounted for only in terms of a holistic approach that appeals to social concepts not reducible to what the individuals are personally committed to do. In contrast, for other authors, like Robert Sugden and Ken Binmore, for instance, beliefs that one ought to conform to the prevailing conventions develop on top of such regularities and recruit natural human sentiments. In this naturalistic perspective, moral norms develop out of mere social conventions.

Luca Tummolini

See also Common Knowledge; Conventions, Logic of; Cooperation/Coordination; Evolutionary Game Theory and Sociality; Social Institutions; Social Norms; Social Ontology, Recent Theories of; Social Rules; Spontaneous Order

Further Readings

Binmore, K. (2005). *Natural justice.* Oxford, England: Oxford University Press.

Gilbert, M. (1992). *On social facts* (2nd ed.). Princeton, NJ: Princeton University Press.

Hume, D. (1740). *A treatise of human nature* (2nd ed.; L. A. Selby-Bigge, Ed.). Oxford, England: Clarendon Press.

Lewis, D. (1969). *Convention: A philosophical study.* Cambridge, MA: Harvard University Press.

Skyrms, B. (1996). *Evolution of the social contract.* Cambridge, England: Cambridge University Press.

Sugden, R. (2004). *Economics of rights, co-operation and welfare* (2nd ed.). Basingstoke, England: Palgrave Macmillan.

Tummolini, L. (Ed.). (2008). Convention: An interdisciplinary study. *Topoi, 27,* 1–164.

Social Epistemology

In simplest terms, social epistemology is the normative study of knowledge as a social product. It is a cross-disciplinary nomad, equally at home in philosophy and policy. There is disagreement over whether it is meant to be a branch of epistemology or sociology, or rather the entirety of one or the other or both of these disciplines. This entry begins by discussing three types of social epistemology that canvass these possibilities, followed by an extended discussion of the most ambitious form of social epistemology, which attempts to bridge the analytic/continental divide within contemporary philosophy, while providing an account of the social construction of intellectual progress.

Three Types of Social Epistemology

Social epistemology may be regarded in one of three ways: (1) as a branch of sociology, (2) as a branch of epistemology, or (3) as a field that transcends the difference between (1) and (2). Let us take each in turn.

1. As a branch of sociology, social epistemology asserts that social relations can be organized in terms of the differential, often hierarchical, access that a society's members have to a common reality. Plato originally advanced a static version of this thesis in the *Republic.* There, each level of human understanding—from the ideal to the base—corresponded to a stratum in a myth-based caste system. A more dynamic version, based on the stages of human intellectual progress, was advanced more than 2,000 years later by Auguste Comte in his positivist polity. In this context, earlier religious and metaphysical forms of epistemic authority served atavistic class–like functions in a science-led social order. The general idea continues to fascinate philosophers, as it raises the prospect of nonviolent, large-scale social control by deference to expertise, also known as "division of cognitive labor." Indeed, such knowledge-based politics is arguably the most Machiavellian of all, as it delegates the application of force to individuals, whose willed compliance is socially rewarded with the assignment of rationality. The history of medicine probably provides the clearest traces of this issue.

2. As a branch of epistemology, social epistemology asserts that an adequate grasp of the state of knowledge in society requires more than generalizing from what a single ideal (Cartesian) or average (Humean) mind knows. It requires recognizing the distributed nature of knowledge, either emergent on specific forms of life (i.e., folkways) or divided according to some overarching rational plan (e.g., science). But in both cases, the whole knowledge system is much more—and even other—than

the sum of what individuals know. On the one hand, theories of rationality and the scientific method—perhaps most strikingly Karl Popper's falsificationism—operate as a corrective social norm to the cognitive liabilities of individuals; on the other, no collectively agreed scientific theory of the sort that might govern a Kuhnian paradigm is likely to be known entirely by anyone or equally well known to everyone who claims adherence. In this respect, the fallacies of composition and division mark the boundary of social epistemology.

3. Finally, *social epistemology* may simply be a good contemporary name for epistemology's original project—put in Hegelian terms, to render knowledge "self-conscious." This point pertains to the English coinage of *epistemology* in the mid-19th century as a distinct branch of philosophy concerned with the conditions under which things may be known, where the very idea of the "unknowable" was treated as an oxymoron symptomatic of a (presumably defunct) metaphysics in which humanity remained permanently alienated from the divine source of reality's intelligibility. To be sure, many—perhaps even most—things remained unknown, but nothing was unknowable. Thus, alongside epistemology was christened *agnoiology,* the field that was supposed to capture the sphere of remaining ignorance in the spirit of an achievable research program. The main secular descendant was the logical-positivist aspiration for a unified scientific language in which anything worth saying could be said. By the 1960s, courtesy of Thomas Kuhn and Michel Foucault, this aspiration had become relativized to successive phases (paradigms or *epistemes*) in the history of science.

The difference between the first and second types of social epistemology is encapsulated in the chart below. The chart highlights the extent to which knowledge can be seen as either (a) a stabilizing or (b) a dynamizing force in society. The third type, to which the rest of the entry is devoted, may be seen as a synthesis of the first and the second.

Social Epistemology as the Goal of All Epistemology

An interesting story yet to be told is how epistemology—a field originally conceived of as a vehicle for achieving "absolute knowledge" (more about this is given below)—acquired a renewed taste for the problem of skepticism once it fell into the hands of post-positivist "analytic" philosophers, who have dominated the discipline in the Anglophone world since the end of World War II. These philosophers have resorted to various "foundational" moves, ranging from introspectively based indubitable intuitions, as in the case of Roderick

Table 1 Social Epistemology as (1) Sociology and (2) Epistemology

Social Organization of Knowledge	*(1) Individually Distributed and Collectively Reproduced*	*(2) Collectively Produced and Individually Redistributed*
Symbol of social epistemology	Evolution (especially irreversible specialization)	Encyclopedia (especially crosscutting referencing)
Knowledge as a principle of social order	Expertise and stratification	Research and education
Division of labor	Fixed and interdependent	Temporary and exploitative
Knowledge vis-à-vis power	Integrative ("social equilibrium")	Disintegrative ("creative destruction")
Sociological exemplar	Stable ecology	Dynamic university
Discovery vis-à-vis justification	Discovery is disparate, and justification channels to the mainstream (tributaries)	Discovery has biased origins, and justification redistributes privilege (delta)

Source: Author.

Chisholm, to more objectively determined reliable processes, as in the case of Alvin Goldman. Although they often cite Bertrand Russell's definition of knowledge as "justified true belief" as the touchstone for their approach, it is clear that Russell's default epistemic position was more positive than that of any of his analytic offspring. Whereas Russell defines knowledge so as to remove a priori restrictions (i.e., biases, prejudices, and blind spots) to our understanding of reality, which he believed the early-20th-century revolutions in mathematical logic and relativity physics had already begun to do, Chisholm and Goldman are more concerned with protecting whatever we know from contamination by error. In this respect, a latter-day concept that more closely captures Russell's original spirit is *epistemic injustice,* which aims to reduce the power effects of knowledge claims by proportioning their credibility to the weight of evidence in their favor.

One explanation for the risk-averse character of analytic philosophy is that two science-led world wars have shifted the default epistemic starting point from "how to go forward" to "how not to slip backward." That backslide in the default position had already begun in the aftermath of World War I, which by 1920 had witnessed a learned obituary on the idea of progress (J. B. Bury's *The Ideal of Progress*) and an ominous prophecy about the future of Western civilization itself (Oswald Spengler's *The Decline of the West*). Against this backdrop, the logical positivists in the following decade sought epistemic foundations based not on natural induction, common sense, or collective memory but on an agreed observation language through which knowledge claims could be verified. Their main concern was to penetrate disciplinary jargons that overstated ("totalized") the jurisdiction of their knowledge claims, reaching into areas of life where individuals were entitled to free choice. An alternative strategy at the time for dealing with this problem, inspired by Max Weber and promoted by Karl Jaspers, aimed to instil rhetorical self-restraint on academic experts by teaching them how their field's specific history has conditioned their worldview. Nowadays, this approach, especially popular in the social sciences, is called "reflexive," a word that Alvin Gouldner introduced to radicalize American sociology's self-understanding, where the generality of the discipline's discourse masked the specificity of its own history and that of the society that had shaped it.

Two other philosophers who came of age in the 1920s, but whose influence would peak only after World War II, Theodor Adorno and Martin Heidegger, proposed still more drastic solutions to the crisis of epistemic legitimacy in the sciences. Whereas Adorno proposed endless self-criticism of the power relations that normally legitimize knowledge claims, Heidegger advanced a philologically inspired strategy to recover the ground of ultimate being in its original Greek manifestation. Between them, the "made for export" market for "continental" European philosophy was defined in the postwar period. The "crisis" mentality reflected a lost sense of organized inquiry as integral to humanity's collective self-realization as a species. What before World War I had been celebrated as an increasingly rationalized division of cognitive labor had come by the end of World War II to be diagnosed as a debilitating fragmentation of inquiry. Concern about the alienating, if not outright dehumanizing, tendencies accompanying the advancement of science was given its most mature and articulate expression in a series of public lectures in the 1930s by the phenomenologist Edmund Husserl, who canonized the distinction between the sciences' rival "systematic" visions of the world and that of a presumptively coherent human *life-world*. The reconciliation of these two visions captured the imagination of postwar philosophers on both sides of the analytic/continental divide, most notably Wilfrid Sellars and Jürgen Habermas.

Social Epistemology as Bridging the Analytic/Continental Philosophy Divide

An argument can be made that social epistemology has been always central to both analytic and continental trends in modern European philosophy, despite their strikingly opposed ways of characterizing the overall social dynamic of knowledge. This contrast is epitomized in their respective signature phrases for embodied social knowledge: *common sense* and *collective memory.* The difference between the two can be explained in terms of their default theological starting points. Common-sense theorists take the reliability of our mental faculties to be underwritten by our divinely created souls, the repository of a priori knowledge, whereas collective memory theorists take those faculties to be born contaminated by our animal bodies, perhaps as living reminders of original sin. To be sure, both positions have been secularized

over the past 200 years: On the one hand, the seat of common sense migrated from a specifically Christian (as in Thomas Reid) to a more generically Platonic (as in G. E. Moore) sense of the soul; on the other (discussed in more detail below), collective memory has lost its original associations with theodicy (Hegel) to embrace more explicitly materialist conceptions of progress (Marx). The social epistemology of science can be more specifically understood as secular variants of both traditions. So-called naturalists and evolutionary epistemologists, exemplified by John Dewey, treat science as an extension of common sense, whereas self-styled critical rationalists, exemplified by Karl Popper, regard science as being in constant struggle against the errors committed in its own history.

Whereas the common-sense theorist values knowledge that enables one to come closer ("correspond") to the world as it naturally is, the collective memory theorist values knowledge that enables one to stand apart ("self-differentiate") from the world as a second-order entity—ideally, to remake the world in one's own image. The one privileges conformity, the other autonomy as an epistemic virtue: that is, getting it right versus thinking for oneself. The two social epistemologies differ interestingly on induction as a source of knowledge. For the common-sense theorist, induction is a positive mental tendency that reflects accumulated experience, whose sheer survival provides a forward momentum for future epistemic judgments. In contrast, the collective memory theorist regards induction more negatively as the path-dependent drag of the past that fails to distinguish what is truly needed for effective future action. What the former regards as wisdom, the latter treats as prejudice. Important strands of 20th-century psychology have tried to reconcile these two images: Carl Jung's brand of psychoanalysis may be seen as having aspired toward a synthesis, while nowadays it is common to adopt the usage of the cognitive psychologist Daniel Kahneman, describing induction as a "bias," that is, neutral to the two positions.

When James Ferrier, a Scottish importer of German idealism, coined *epistemology* in 1854, he was operating from a collective memory perspective, in explicit contrast to the common-sense tradition dominant in his homeland. Ferrier defined knowledge as a second-order awareness of our mental states, a conscious organization of experience into a systematic whole—recalling that the Cambridge Christian Platonist Ralph Cudworth had introduced "consciousness" into English in 1678 as the seat of personal identity (aka soul), on the basis of which we judge in this life and are judged in the next one. In other words, the mind always already contains a representation of reality's validation mechanism that enables it to recognize the truth even when it breaks with what is expected or desired, a capacity familiar in ethical contexts as "conscience." In the three centuries after Cudworth, this idea was divested first of its theological baggage and then, in the 20th century, even of its psychological baggage. Thus, with the advent of logical positivism, the philosophical logician Alfred Tarski secularized this second-order awareness as a *metalanguage,* or, more simply, *semantics,* that provides the truth conditions for our beliefs, now understood as first-order statements. By the 1970s, thanks largely to Saul Kripke, this idea became the cornerstone of the theory of reference, specifically the concept of *semantic reference.*

In terms of this overall trajectory, Ferrier is a transitional figure whose nod to theology remains in his definition of the goal of knowledge as the *absolute,* which alludes to the Christian idea of absolution from sin. Relevant here is Hegel's background in theology, especially his modeling of the progress of the world-historic spirit as theodicy played out on a temporal stage: that is, evils that God always already justifies from the standpoint of eternity, which we will come to terms with only in the fullness of time. The corresponding vision of the human knower is as an agent who proceeds through the world with, so to speak, a "dirty mind," just as, morally speaking, we proceed with "dirty hands," both of which have the capacity to become "cleaner" over time. More prosaically, epistemology is about separating the wheat from the chaff of our beliefs and integrating the former so that a larger truth emerges that gives direction to our inquiries and provides meaning to our lives. It is what Marxists call the *dialectical method.* Just as sin is inevitable in a world where fallible beings are called to act, so too is error inevitable in a world where we cannot learn anything at all without relying on our fallible senses. In this respect, the Popperian imperative to test our spontaneously formed beliefs is the epistemic equivalent of a moral conscience.

Left hanging in the balance, even in these post-theological times, is the legacy of Hegel's treatment of absolution as a self-reflexively applied process—that is, humanity's collective capacity to absolve

itself of sin/error without explicit divine guidance. This idea has proved a source of hope for ambitious politicians over the past two centuries, most notably Marxists, who have desired to build a "Heaven on Earth." But the view has also had its admirers among those more clearly associated with the canon of scientific epistemology: for example, the pragmatist Charles Sanders Peirce's self-correcting version of scientific inquiry, understood as a collective process that over time "converges" on the ultimate representation of reality—absolute truth. To be sure, Peirce's conception is ambiguous with regard to the nature of validation: Does a maximally comprehensive scientific consensus emerge from a common recognition of the truth, or is the truth itself the product of such a consensus? Put theologically, are we validated *by* or *as* God? The logic of this idea has been explored interestingly within analytic social epistemology by Frederick Will, yet it remains largely neglected by the field's practitioners.

Perhaps the issue on which analytic social epistemology comes closest to a continental sensibility is the distinction in science's contexts of *discovery* and *justification,* which presupposes that truth can be extracted from history to become part of an ongoing collective inquiry. Nevertheless, from a continental standpoint, analytic philosophers show remarkably little concern for exactly how this extraction is supposed to take place in practice. Instead, they have emphasized the output of such an extraction—namely, a canonical representation of a new discovery's place within an established theoretical framework. The overall rhetorical effect is to suggest that a discovery could have been made by other inquirers by other means—such that, say, the fact that Isaac Newton discovered the laws of motion or Charles Darwin the theory of natural selection is incidental to the validity of what they discovered—and that others with rather different training and interests may extend or apply these discoveries in the future. Indeed, such is the process by which science becomes "universal knowledge" in a sense that Hegel would have recognized.

Conclusion: Social Epistemology and the Social Construction of Progress

Analytic social epistemologists such as Alvin Goldman and Philip Kitcher act as if a rational future lies in people coming to the realization that they know more and more about less and less, which will lead them to cede more of their epistemic authority to others. Perhaps Auguste Comte would approve of this trajectory, but the implied epistemic ecology is not sustainable. For even if people are naturally inclined to defer to authority, the funding costs of indulging this inclination will become prohibitive. Put crudely, if we think that we need increasingly specialized research to solve persistent social problems, then a problem in the conceptualization of these problems is implicated—and the best place to locate the source of this second-order problem is the "politics of science." After all, an all too easy way for scientists to justify the need for more research on their own terms is simply to argue that we do not know enough to make sensible policy decisions. Yet, in most cases, we already can make perfectly decent policy decisions—provided that institutional safeguards are in place to enable those decisions to be reversed if they result in more harm than good. In this context, scientists can strategically enhance their employment prospects by playing on the perceived untrustworthiness of politicians as well as any latent aversion to risk in the public at large.

Nevertheless, as Randall Collins has demonstrated in an exhaustive cross-cultural history of institutionalized intellectual life, the levels of conceptual sophistication and empirical breadth demanded of philosophical and scientific inquiry have been subject to ebbs and flows depending on the ambient political economy. There is no unequivocal measure—other than sheer quantity of output—in terms of which it can be said that humanity, or even just the West over the past 200 years, has made progress. However, what is clear is that any expansion or contraction of the discursive space for knowledge production can be—and has been—justified as progressive. Thus, a growing research environment tends to be seen more in terms of opening up new horizons than as exhibiting a dispersion of effort, whereas a shrinking environment gets interpreted as consolidating and focusing effort instead of arresting development. Here, social epistemology could make greater use of Leon Festinger's cognitive dissonance theory of social psychology, especially its key concept of "adaptive preference formation," whereby thwarted expectations provide an opportunity for people to rethink their priorities, in the course of which they reorganize their relationships to the past so as to provide the legitimatory grounds

for embarking on a different future. Indeed, as Kuhn originally observed, this is one of the most important yet still least remarked achievements of scientific textbooks in the aftermath of a scientific revolution.

Steve Fuller

See also Collective Memory; Common Sense (in the Social Sciences); Disagreement; Encyclopedia; Epistemology; Feminist Epistemology; Objectivity; Philosophy of Expertise; Social Studies of Science and Technology; Sociology of Knowledge and Science; Strong Program in the Sociology of Scientific Knowledge; Virtue Epistemology

Further Readings

Collins, R. (1998). *The sociology of philosophies*. Cambridge, MA: Harvard University Press.

Foucault, M. (1970). *The order of things*. New York, NY: Random House. (Original work published 1966)

Fuller, S. (1988). *Social epistemology*. Bloomington: Indiana University Press.

Fuller, S. (2003). *Kuhn vs. Popper: The struggle for the soul of science*. Cambridge, England: Icon.

Fuller, S. (2007). *The knowledge book: Key concepts in philosophy, science and culture*. Durham, England: Acumen.

Fuller, S. (2011). Science and technology studies and social epistemology: The struggle for normativity in social theories of knowledge. In I. Jarvie & J. Zamora-Bonilla (Eds.), *The SAGE handbook of philosophy of social sciences* (pp. 665–685). London, England: Sage.

Fuller, S., & Collier, J. (2004). *Philosophy, rhetoric and the end of knowledge*. Mahwah, NJ: Lawrence Erlbaum. (Original work published 1993 by S. Fuller)

Goldman, A. (1999). *Knowledge in a social world*. Oxford, England: Oxford University Press.

Goldman, A., & Whitcomb, D. (Eds.). (2011). *Social epistemology: Essential readings,* New York, NY: Oxford University Press.

Gouldner, A. (1970). *The coming crisis in Western sociology*. New York, NY: Basic Books.

Haddock, A., Miller, A., & Pritchard, D. (Eds.). (2010). *Social epistemology*. Oxford, England: Oxford University Press.

Husserl, E. (1954). *The crisis of European sciences and transcendental phenomenology*. Evanston, IL: Northwestern University Press. (Original work published 1937)

Kitcher, P. (2001). *Science, truth and democracy*. Oxford, England: Oxford University Press.

Kuhn, T. (1970). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Social Facts

Émile Durkheim introduced the concept of "social facts" at the end of the 19th century to designate those forms of social thought, feeling, and action that form the distinctive subject matter of sociology. This entry considers the accounts of social facts offered by early social theorists such as Durkheim and Max Weber, notes the general neglect of the concept of sociality in later 20th-century sociology and philosophy of social science, and briefly describes the recent resurgence of philosophical interest in the analysis of social facts.

Origins

Émile Durkheim, one of the founding fathers of sociological science, claimed in the *Rules of Sociological Method* that it was important to identify the distinctive characteristics of social phenomena in order to distinguish the subject matter of the new scientific discipline of *sociology* from the subject matters of established sciences such as psychology and biology. He noted that it is not enough to characterize social facts as facts that are generally distributed in society, since many psychological and biological facts, such as reasoning and drinking, are also generally distributed in society.

Durkheim defined social facts as forms of thought, feeling, and action that are *external* to individuals and that exert a *coercive* force upon them. Unfortunately, in his zeal to validate the scientific status of the new discipline of sociology, Durkheim cited characteristics that are distinctive of the objects of scientific inquiry rather than sociality per se (hydrochloric acid and gravity are also external to individuals and exert a constraining force upon them, but they are not social in nature). Durkheim also treated social facts as statistical facts about social groups, such as the different rates of suicide between Catholics and Protestants documented in his monograph on *Suicide*, but this presupposes some independent account of social groups. However, given the examples of social facts that Durkheim discussed, such as family obligations and religious practices, it is clear that he conceived of

social forms of thought, feeling, and action as forms of human psychology and behavior normatively oriented to the represented psychology and behavior of members of social groups.

In *Economy and Society,* Max Weber, another founding father of sociology, identified its subject matter as social action, defined as meaningful behavior that takes into account the behavior of others and is oriented in its course. Weber is often misrepresented as equating social action with interpersonal action, which would include doubtful social actions such as acts of aggression and rape. Yet his own illustrative examples indicate that he considered action to be social only if it is *normatively* orientated to the represented behavior of members of social groups. Accordingly, like Durkheim, Weber denied that merely imitated actions, such as copying another's superior method of baiting a fishing line, or merely common actions, such as a crowd of people raising their umbrellas to avoid rain, are social actions. Or, as Durkheim put it, social facts are imitated or common because they are social, not social because they are imitated or common.

At the root of Durkheim and Weber's conceptions of sociality is the notion of an individual's psychological connection to other individuals represented as members of a social group, and the normative orientation of an individual's psychology and behavior to the represented psychology and behavior of members of that social group. Thus, both Durkheim and Weber were careful to distinguish between genuinely social groups, within which there is a psychological connection and a normative orientation, such as Catholics and teenage gangs, and mere aggregate groups of individuals who happen to share a common property or set of properties, such as the populations of persons who are ambidextrous or are afraid of spiders.

Durkheim famously affirmed the *supra-individuality* of social groups—the notion that social groups have properties over and above the properties of the individuals who compose them—and Weber famously denied it, leading to their historical representation as champions of *holism* and *individualism,* respectively. However, both in practice treated social groups as virtual realities, as populations of individuals representing themselves as members of groups normatively bound by shared forms of thought, feeling, and action, a view most closely associated with the sociologist Georg Simmel and the Gestalt social psychologist Solomon Asch.

While many early-20th-century social scientists did embrace this conception of sociality, later sociologists abandoned foundational questions about social facts as they focused on developing substantive *structural, functional,* and *structural-functional* explanations of the diverse phenomena they designated as social, while later social psychologists focused almost exclusively on interpersonal interaction and cognition. For many years, philosophers of social science also ignored foundational questions by equating social action with meaningful action and endlessly debating the question of whether meaningful human action could be causally explained.

Recent Developments

In recent decades, however, some philosophers have returned to foundational questions about the nature of social facts. This new focus on sociality was stimulated in large part by Margaret Gilbert's pioneering work *On Social Facts,* in which she explicates the notion of psychological connectedness between members of social groups in terms of populations of individuals conceiving of themselves as plural subjects, bound by joint commitments to beliefs, actions, and the like. Similar accounts have been developed by John Searle, Michael Bratman, and Raimo Tuomela in terms of collective intentionality, shared intentionality, and joint acceptance, respectively, although to date, these analyses have had little impact on mainstream social science compared with their influence on recent trends in the philosophy of social science.

John D. Greenwood

See also Collective Agents; Collective Intentionality; Durkheim's Philosophy of Social Science; Group Identity; Holism, in the Social Sciences; Individualism, Methodological; Plural Subjects; Searle and the Construction of Social Reality; Social Ontology, Recent Theories of; Structural Functionalism, in Social Theory; Weber's *Verstehende* Approach; We-Mode, Tuomela's Theory of

Further Readings

Durkheim, É. (1895). The rules of sociological method. In S. Lukes (Ed.), *Durkheim: The rules of sociological method and selected texts in sociology* (W. D. Halls, Trans., pp. 1–165). New York, NY: Macmillan.

Gilbert, M. (1989). *On social facts.* London, England: Routledge.

Greenwood, J. D. (2003). Social facts, social groups and social explanation. *Nous, 37,* 93–112.
Tuomela, R. (2003). Collective acceptance, social institutions, and social reality. *American Journal of Economics and Sociology, 62*(1), 123–165.
Weber, M. (1922). *Economy and society* (2 vols.; G. Roth & E. C. Wittich, Eds.). Berkeley: University of California Press.

Social Institutions

The entry focuses on a central feature of social life: social institutions. It explains their core features and their types, presents some of the theories of how they arise, and brings to our attention points of contact as well as points of divergence between social institutions, on the one hand, and rules, norms, conventions, and practices, on the other.

A social institution is a normative social structure that influences the behavior of those who participate in it. Institutions facilitate cooperation, and many institutions promote social order. Examples of social institutions are organizations such as the United Nations, the Roman Catholic Church, or the Chinese Communist Party. These structures involve people with particular tasks and roles. Social institutions that are not organizations encompass money, property, and the rules of the road. Institutions are usually closely intertwined with one another. Money and property, for instance, cannot function properly without banks and notaries. The study of social institutions plays a central role in a number of social sciences, including cultural anthropology, economics, political science, and sociology. Social scientists analyze institutions frequently as patterns in behavior or as social rules. They are social constructs that often evolve spontaneously.

Activities and Rules

The institution of money involves a number of characteristic activities, including exchanging cash for goods at a cash register. The regular performance of such an action by a number of people establishes a behavioral pattern or social practice. Institutions are sometimes identified with behavioral patterns. Such analyses leave out the normative dimension of institutions, which differentiates them from social practices. Rules are normative, and they tend to issue in the behavior they regulate. So rules account for both the normative and the behavioral dimensions of institutions. A rule is a social rule when a number of people share the belief that it is in force. Institutions are social because they consist of social rules.

Institutions cannot exist without intentional attitudes such as beliefs, desires, and intentions. They are sometimes identified with (shared) beliefs about which behavior is appropriate, permitted, required, or forbidden in a particular kind of situation, that is, with (interlocking) normative expectations. A customer would be surprised if the item she wants to buy is put behind the counter while the cashier puts the money that she hands to him in his own wallet. Her (presumably shared) expectation would be violated. There might not be more to institutions than such shared normative expectations. Note that the content of a normative expectation is a rule and that a set of interlocking normative expectations is a structure of expectations. So the claim at hand is consistent with the thesis that institutions are social rules and that they are normative social structures that affect behavior.

Social norms are social rules, while conventions are frequently explicated in terms of behavioral regularities. Institutions are similar but not identical to social norms and conventions. Suppose conventions are patterns in behavior that are arbitrary in some sense (as in David Lewis's work). Then driving on the left side of the road is a convention in this sense. This regularity need not have a normative status. So it is not the case that all conventions are social institutions. Furthermore, a rule that issues in a certain behavioral regularity can be an institution without being arbitrary. There is nothing arbitrary about being obliged to do as promised. Hence, not all institutions are conventions. It may be that all social norms are institutions, even though not all institutions are social norms. Social norms might be general in the sense that they apply to anyone in a particular situation, for instance, to anyone who has made a promise. Institutional roles apply only to those who occupy the relevant role. Only the president, for instance, has the right to veto a law. Note that institutions do not only constrain but also enable actions. Without money, for instance, we would still be operating in a barter economy.

Spontaneous Evolution and Social Construction

Institutions in general and economic institutions in particular are often said to evolve spontaneously

in the sense that they are the unintended consequences of individual actions (Friedrich Hayek is the prime exponent of this thesis). They arise and survive because they facilitate cooperation and are advantageous to their participants. This is captured by models that represent institutions as solutions to social interaction problems (game-theoretic equilibria). Such spontaneous evolution models explain how particular institutions have evolved. Many institutions, however, are deliberately created or designed. It is not obvious that spontaneous evolution models are useful for understanding those features of institutions that were deliberately created. It is sometimes argued that such a model explains how a deliberately created institution *might* arise or why it remains in existence.

Institutions are social constructs in more than one sense. They are socially caused or produced. Furthermore, they are socially constituted in the sense that an institution exists only if it is agreed or collectively accepted to exist. Money, for instance, is a social construction in that certain items are money (if and) only if they are collectively accepted to be money. John Searle argues that the very existence of institutions is to be accounted for in terms of constitutive rules. The structure of constitutive rules, on his theory, is "X counts as Y in context C." Certain pieces of metal (X) can, for instance, count as money (Y) in a particular country (C). The notion of a constitutive rule is useful for explicating how people confer an institutional status on an action, person, or object when they collectively accept that the relevant entity has that status (by virtue of which, some would say, certain regulative rules apply to it). Searle's "counts-as" locution captures the constitutive sense in which institutions are socially constructed.

Frank Hindriks

See also Evolutionary Game Theory and Sociality; Institutionalism and Institutional Theory; Institutions as Moral Persons; Performative Theory of Institutions; Searle and the Construction of Social Reality; Social Conventions; Social Norms; Social Ontology, Recent Theories of; Social Practices; Social Rules; Spontaneous Order

Further Readings

Hayek, F. A. (1988). *The fatal conceit: The errors of socialism* (W. W. Barley III, Ed.). Chicago, IL: University of Chicago Press.

Lewis, D. (1969). *Convention: A philosophical study.* Cambridge, MA: Harvard University Press.

Searle, J. R. (2010). *Making the social world: The structure of human civilization.* Oxford, England: Oxford University Press.

Tuomela, R. (2007). *The philosophy of sociality: The shared point of view.* Oxford, England: Oxford University Press.

Social Interactions: Individual Decision and Group Formation

This entry introduces the growing field of social interactions, created by economists and econometricians to study the effects of peer groups on individuals' decisions. The theoretical and empirical advances in this field help explain a number of otherwise puzzling social phenomena and offer empirical support to the influence of group actions and beliefs on individuals' choice. This novel approach to the study of social phenomena and individual action is an important development that enriches the philosophy of the social sciences, while at the same time showing the significance of the interconnections among social sciences.

Background

The study of social interactions phenomena starts roughly with Gary Becker's celebrated 1974 article on social interactions in the *Journal of Political Economy* but has received prominence since Charles Manski's pathbreaking elucidation, in the *Review of Economic Studies* in 1993, of the pitfalls of quantitative analysis in that area. Additional major contributions since then by William Brock and Steven Durlauf have succeeded in building new links with social capital and with the statistical mechanics, complexity, and emergence literatures.

The decisions of individuals who share a social milieu are likely to be interdependent. Recognizing the exact nature of such interdependence in various settings and measuring empirically the contribution of social interactions to observed choices poses complex methodological questions. Resolving such questions can be critical for a multitude of phenomena in economic and social life and for public policy. The variations in crime rates across cities in the same country and what can be done about influencing

them, the emergence and persistence of residential segregation, and its variation across different countries are all in principle issues that may be examined as social interactions phenomena.

Individuals care about private outcomes and consumption of private goods; they may value how their own consumption of goods influences others or care about the kinds of cars others drive, the incomes of their friends, or the education acquired by their friends' children. The latter type of effect is an interpersonal effect, known as *endogenous* social effect (or interaction), because it depends on *decisions* of others in the same social milieu. Individuals may also care about the personal and demographic characteristics of others, and the associated effects are known as exogenous social or *contextual* effects. In addition, individuals in the same or similar individual and social settings tend to act like one another because of factors they share in common. Such an interaction pattern is known as *correlated* effects. These canonical distinctions, due to Charles Manski's 1993 paper on the "social reflection problem," have helped streamline the literature.

Distinguishing empirically between endogenous, contextual, and correlated effects can be critical for policy analysis, as explained further below, and has drawn considerable attention in the literature.

When individuals belonging to the same social group are influenced by the actions of others, even if they take those actions as given, their actions are interdependent. The actions of all group members are determined simultaneously, and the actions of each group member depend on her own as well as everyone else's characteristics. The theoretical and empirical tools that economists have developed break this possibly infinite regress.

By choosing where to live or what group to belong to, individuals choose in effect their neighborhood effects, or their social context, more generally. Such choices involve underlying factors that may be in part unobservable to the analyst and, therefore, require making inferences among the possible factors that contribute to decisions.

Analytical Framework

Let individual i's decision, say years of education, ω_i //I////I//, be a linear function of a vector of observable individual characteristics, X_i, of a vector of contextual effects, $Y_{n(i)}$, which describe i's neighborhood, $n(i)$, and of the expected value of the ω_j of the members of neighborhood $n(i)$, $j \in n(i)$, that is, the expected years of education of all others in the neighborhood. It is straightforward to incorporate social interactions into economic models in a manner that is fully compatible with economic reasoning and obtain a behavioral equation such as

$$\omega_i = k + cX_i + dY_{n(i)} + Jm_{n(i)} + \varepsilon_i, \tag{1}$$

where ε_i is a random error and k a constant. Ignore for the moment the fact that individual i may have deliberately chosen neighborhood $n(i)$. The critical next step is to assume *social equilibrium* and that individuals hold *rational expectations* over $m_{n(i)}$. So taking the expectation of ω_i and setting it equal to $m_{n(i)}$ allows us to solve for $m_{n(i)}$. Substituting back into Equation 1 yields an expression for individual i's outcome in terms of all observables:

$$\omega_i = \frac{k}{1-J} + cX_i + \frac{J}{1-J}cX_{n(i)} + \frac{d}{1-J}Y_{n(i)} + \varepsilon_i. \tag{2}$$

This simple linear model confirms that endogenous social effects generate feedbacks that magnify the effects of neighborhood characteristics. That is, the effect of a unit increase in $Y_{n(i)}$ is $d/(1-J)$, from Equation 2, and not just d, as one would expect from Equation 1.

Identification

Charles Manski was the first to emphasize that the practice of including neighborhood averages of individual effects as contextual effects, $Y_{n(i)} = X_{n(i)}$, may cause failure of identification of endogenous as distinct from exogenous interactions, that is, estimating J separately from d. That is, if the neighborhood attributes *coincide* with the neighborhood averages of its inhabitants' characteristics, or $Y_{n(i)} = X_{n(i)}$, then regressing individual outcomes on neighborhood averages of individual characteristics as contextual effects allows us to estimate a function of the parameters of interest, $(Jc + d)/(1-J)$, the coefficient of $X_{n(i)}$ in a regression according to Equation 2. A statistically significant estimate of this coefficient implies

that at least one type of social interaction is present, that is, either J or d, or both are nonzero. This is known as Manski's *reflection problem,* which is specific to *linear* models: The equilibrium value of the outcome $m_{n(i)}$ is linearly related to the neighborhood attributes entering the causal model, and therefore its effect on individual outcomes may not be distinguishable from their "reflection."

Complicating the basic model in natural ways, as by assuming correlated effects—the performance of different students in the same class is affected by the quality of their teacher, which is unobservable, over and above peer effects from classmates—introduces additional difficulties with identification, even if individuals are randomly assigned to groups.

If it is plausible to exclude some neighborhood averages of individual covariates from the causal model, then identification may be possible. Also, if nonlinearities are inherent in the basic-model specification, identification again may be possible, even in the case where the contextual effects coincide with the neighborhood averages of individual characteristics. An interesting consequence of endogenous social interactions is the amplification of differences in average neighborhood behavior across neighborhoods. In their 2003 study, Edward Glaeser, Bruce Sacerdote, and José A. Scheinkman directly used such patterns in the data to estimate a *social multiplier.* This is defined, for a particular fundamental determinant of an outcome, as the ratio of the total effect of a change in its value—which includes a direct effect on an individual outcome plus the sum total of the indirect effects through the feedback from the effects on others in the social group—to the direct effect. From Equation 1,

$$\frac{d}{1-J}\,\frac{1}{d} = \frac{1}{1-J}\,. \tag{1}$$

This social multiplier is particularly useful in delivering ranges of estimates for the endogenous social effect and when individual data are hard to obtain.

Phenomena like diffusion of innovations, herding and adoption of norms or other institutions, and transmission of job-related information are conceptually related to social interactions. They can be modeled by modifications of Equation 1.

Empirical Findings of the Effects of Social Interactions

Several researchers have sought to identify social interactions by exploiting uniquely suitable features of observational data, such as so-called natural experiments. For example, Bruce Sacerdote, in his 2001 article in the *Quarterly Journal of Economics,* exploits the fact that at Dartmouth College, freshman-year roommates and dorm-mates are randomly assigned. Sacerdote applies Equation 1 and posits that an individual's grade point average is a function of an individual's own academic ability prior to college entrance (an own lagged effect), of social habits, and of the academic ability and grade point average of his roommates. He finds that peers have significant impacts on each others' grade point average and on decisions to join social groups such as fraternities. It is notable that strong peer effects in student outcomes are present even among highly selected college students in close proximity to one another, even though they may be otherwise quite homogeneous.

Education policy and the functioning of schools depends critically on whether students benefit from classmates with different characteristics and academic performance and whether the effect is different depending on whether one's classroom peers are more or less able. This is critical for deciding whether or not students should be "tracked"—that is, administratively segregated in terms of different performance. It is thus interesting to establish whether peer effects exist in classrooms and schools.

The demographic composition of classrooms matters. For example, Peter Arcidiacono and Sean Nicholson reported in the *Journal of Public Economics* in 2003, using data on all students admitted to U.S. medical schools in a particular year, that female students benefit from attending medical schools that have other female students with relatively high scores on the verbal reasoning section of the Medical College Admission Test.

Randomized field experiments, like the Moving to Opportunity (MTO) conducted by the U.S. Department of Housing and Urban Development in several large U.S. cities, have also been used to estimate social interaction effects. The experiments offered poor households (chosen by lottery from among residents of high-poverty public housing projects) housing vouchers and logistical assistance

through nongovernmental organizations for the purpose of relocating to precisely defined "better" neighborhoods. Several studies based on data from these experiments show that outcomes after relocation improved for children, primarily for females, in terms of education, risky behavior, and physical health, but the effects on male youth were adverse.

Nonrandom Sorting

The presence of nonrandom sorting due to factors that are inherently unobservable is a major challenge for the econometric identification of social interactions models. Several papers by Brock and Durlauf showed that this can be an advantage by recognizing that self-selection itself by being a decision—that is, that individuals choose their neighborhoods—makes $n(i)$ in Equation 1 endogenous. Estimating a neighborhood selection rule and correcting for selection bias by means of the so-called Heckman correction term introduces an additional regressor in the right-hand side of Equation 1, whose neighborhood average is not a causal effect. In a 2008 article in the *Journal of Urban Economics,* Yannis Ioannides and Jeffrey Zabel reported results from implementing this method, using micro data for a sample of households and their 10 closest residential neighbors from the American Housing Survey and contextual information for the census tracts in which those households reside. They report a very significant and large endogenous social effect in housing demands—individuals care about the housing consumption of their neighbors—along with very significant contextual effects in the form of unobservable group effects. Several other studies have sought to use instrumental variables to account for self-selection. Still, the identification of valid instruments is often quite hard and requires deep understanding of the actual setting.

Philosophical Issues

Recognition that social effects on individuals' decisions matter brings up a number of philosophical considerations. One is whether or not it matters for free will if individuals are influenced by the actions of others and if they have chosen to be influenced by them versus if they have not. Freedom of association leads to the deliberate formation of social groups and communities. This helps individuals benefit from associating with certain other individuals with desirable attributes and helps distance themselves from others with undesirable attributes. The existence of, say, conformism (or nonconformism) within social groups is an outcome of choice that underpins group formation. A second philosophical consideration is the possibility that social interactions are emergent properties in that they apply at an aggregate level but lack individual analogs, and therefore they are not describable at the individual level. When social interactions may be understood as emergent properties of social groups, delicate issues of causality may be raised. Causality in the presence of emergence is not well understood and deserves attention in future research.

Conclusions

Social interactions are ubiquitous, and estimating their effects has motivated important methodological advances in theory and econometrics. With ever-improving data availability, social interactions empirics will rely increasingly critically on careful theorizing that involves precise definitions of social interactions and also utilizes tools and results from psychology and sociology to define appropriate boundaries and to exploit data from different sources.

Yannis M. Ioannides

See also Collective Rationality; Complexity and the Social Sciences; Decision Theory; Econophysics; Emergence; Emergence and Social Collectivism; Group Beliefs; Mutual Beliefs; Rational Expectations; Social Capital; Social Norms

Further Readings

Blume, L. E., Brock, W. A., Durlauf, S. N., & Ioannides, Y. M. (2011). Identification of social interactions. In J. Benhabib, A. Bisin, & M. O. Jackson (Eds.), *Handbook of social economics* (Vol. 1A, pp. 853–964). Amsterdam, Netherlands: North-Holland.

Brock, W. A., & Durlauf, S. N. (2001). Interaction-based models. In J. Heckman & E. Leamer (Eds.), *Handbook of econometrics* (Vol. 5, pp. 3297–3380). Amsterdam, Netherlands: North-Holland.

Durlauf, S. N. (2004). Neighbourhood effects. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of urban and regional economics: Vol. 4. Cities and geography* (pp. 2173–2242). Amsterdam, Netherlands: North-Holland.

Glaeser, E. L., Sacerdote, B., & Scheinkman, J. (2003). The social multiplier. *Journal of the European Economic Association, 1,* 345–353.

Ioannides, Y. M. (2008). Emergence. In S. N. Durlauf & L. E. Blume (Eds.), *The new Palgrave dictionary of economics* [Electronic version]. New York, NY: Palgrave Macmillan. http://www.dictionaryofeconomics.com/article?id=pde2008_E000313

Ioannides, Y. M. (2012). *From neighborhoods to nations: The economics of social interactions.* Princeton, NJ: Princeton University Press.

Ioannides, Y. M., & Loury, L. D. (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature, 42,* 1056–1093.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies, 60,* 531–542.

Moffitt, R. (2001). Policy interventions, low-level equilibria and social interactions. In S. N. Durlauf & H. P. Young (Eds.), *Social dynamics* (pp. 45–82). Cambridge: MIT Press.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics, 116,* 681–704.

Sobel, M. E. (2006). Spatial concentration and social stratification: Does the clustering of disadvantage "beget" bad outcomes? In S. Bowles, S. N. Durlauf, & K. Hoff (Eds.), *Poverty traps* (pp. 204–229). New York, NY: Russell Sage Foundation.

Social Networks

As with other social-scientific terms, there is a degree of definitional ambiguity, debate, and dissonance about the exact meaning of "social networks." The term *social networks* refers to individuals linked to each other by specified social relations. Social networks are of interest because they imply that social connections among people do matter and, accordingly, have to be studied by social science. For instance, while "personalized exchange among many agents" may encompass the idea of a network, at a more general level, *networks* are understood as visualizations of association that aim at mapping social relations and eliciting social structures and their properties. Accordingly, social networks have been applied not only as analytical devices but also, more descriptively, as ways of organizing relations among agents. This is being done in increasingly complex as well as interesting ways.

The notion of social networks has been, explicitly or implicitly, one of sociology's foremost concepts. Starting with classical sociology (the work of Émile Durkhein and Georg Simmel) and drawing on similar interests in social anthropology and social psychology, the concern with social connections and social interaction to which social networks refer has remained unbroken. Social "networks" draw from a metaphor that has to do with the imagery of "fabric." A fabric owes its unitary appearance and properties to the tightly interwoven threads that constitute it. The metaphor extends this idea to social relations among humans: The invisible but nonetheless real and meaningful social relations and social interactions, in threadlike fashion, are interwoven in the form of social networks. They may form smaller or larger groups in variable ways and manage to structure them. In this sense, social networks are but a shortcut for visualizing and conceptualizing social relations. Obviously, social networks differ fundamentally from other kinds of networks, whether these are technological in character, for instance, the Internet, which has a network-type architecture and operation, or biological, for example, the circulatory and nervous systems.

The study of social networks, with its beginnings as a useful metaphor, has developed into a recognizable subfield with its own specialized terms, methods, and literature. As examples of specialist terminology, it is worth mentioning the use of concepts such as the *ego* (meaning the focal individual) and the *node* (meaning intersections or points of contact that link [*tie*] two or more egos together by one or several types of *social relations*). In addition, terms and notions such as those of *point centrality* (local, global, and betweenness), *transitivity, homophily, bridge, clique,* and several others have been developed and deployed to describe the state of affairs within small-scale and large-scale social networks and the relations and influences exerted within and between them. Thus, for example, homophily, which refers to the tendency to associate with others who share similar traits, has been used to indicate similarity that has consequences in terms of the connections that it helps breed. It inheres in network ties of every kind, as in marriage or friendship, and renders them homogeneous in terms of

several of their features, as Miller McPherson and colleagues have shown.

Furthermore, to facilitate the study of social networks in the handling of relational data, several statistical packages have been developed, such as UNINET, STRUCTURE, and PAJEK. These form part of the trend toward mathematized structural analysis of social relations (something that not everyone welcomes).

Social Capital, Trust, and Network Society

One area in which nonmathematized social networks exert their influence has been that of social capital. According to Robert Putnam, social capital requires a continuous effort of "sociability" and continuously repeated contacts during which mutual recognition by the group members is confirmed in order to sustain the group's cohesion. Thus, social capital and its formation are perceived to be based on social networks. Social capital is involved in the density of ongoing and permanent social relations within social networks and in the prevalence, more or less conscious, of a trusting environment and relations, which themselves are generated by functioning institutions. Besides social capital, trust, too, is operative in analyses of social networks: Trust is conceived as something that operates as a functional prerequisite for the unfolding of social life organized within networks, in the absence of which it would not ensue. Thus, the "network form" emerges as quintessential for continuing human social interaction.

Social networks are the bedrock of another idea that has been advanced by Manuel Castells, namely, that of a "network society." Castells links the social network character of human social activity and organization with the network structure of many of the contemporary digital networking technologies that have become available over the past 25 years. The ensuing human–technological complex empowers social networks in new and unanticipated ways; it allows the overcoming of the limitations of older forms of network organization by expanding the horizon and enabling a variety of reconfigurations that transcend nation–state borders, thereby making possible the globalizing and global expansion of this network society. Hence, Castells claims that the human–technological interface now sets the tone of a new, evolving kind of society. This view has been criticized, however, as one that invites the fallacy of technological determinism.

Social Networks and Economic Sociology

A renewed interest in social networks has surfaced with the re-emergence of economic sociology since the 1970s. Social networks and their analysis have been utilized widely in studies as thematically varied as those on business and work organizations, processes of innovation, markets (in particular labor markets), entrepreneurship, interfirm relations, and socioeconomic development, and in eliciting the role of social capital and trust for economic purposes. Nor has the level of analysis been unitary. Consequently, it may assume diverse forms, such as analysis at the level of the individual actor around whom the network revolves (actor centrality), that of the networks' structure (systemic), or that concerned with the flows within social networks.

An example of the use of social network analysis in teasing out regular structures of social ties, and hence social-structural patterns of social behavior, is found in the work of Mark Granovetter, who studied patterns of job searching. Subjects were asked how they managed to obtain their regular job and were required to list a number of channels through which they obtained the relevant information for applying for the job, such as recruiting agencies, official announcements, ads in the press, and so on. It became apparent, however, that a considerable number of interviewees mentioned obtaining the relevant job-related information from a person who was close to them—a "strong tie" in social network terminology. Hearing about the job through an intermediary—a person little known to them, a "friend of a friend," in other words, a "weak tie"—was another important way to elicit such information. Accordingly, the social networks into which one is embedded offer opportunities but also may block them by imposing constraints. Thus, job seekers of low socioeconomic status (SES) cannot utilize their strong or weak ties, which also are low SES, to identify and safeguard a high-level job; here, social class acts as a filter or barrier.

On the whole, social networks that express patterns of social interconnection and interaction imply that social structures—repeated social

practices—take precedence over the individual, exerting not an absolute but still a constraining influence.

Sokratis Koniordos

See also Complex Networks Theory and Social Phenomena; Economic Sociology; Social Capital; Trust, Social

Further Readings

Castells, M. (2010). *The information age: Economy, society and culture. The rise of the network society* (Vol. 1, 2nd ed., with a new preface). Oxford, England: Blackwell.

Degenne, A., & Forse, M. (1999). *Introducing social networks* (A. Borges, Trans.). London, England: Sage.

Freeman, L. C. (2004). *The development of social network analysis: A study in the sociology of science.* Vancouver, British Columbia, Canada: SP Empirical Press.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360–1380.

Koniordos, S. M. (Ed.). (2005). *Networks, trust and social capital: Theoretical and empirical investigations from Europe.* Aldershot, England: Ashgate.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27,* 415–444.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community.* New York, NY: Touchstone.

Scott, J. (2005). *Social network analysis: A handbook* (2nd ed.). London, England: Sage.

Wasserman, S., & Faust, K. (Eds.). (1994). *Social network analysis: Methods and applications.* Cambridge, England: Cambridge University Press.

Social Neuroscience

Neuroscience refers to the collection of disciplines concerned with the structure and function of the brain and, more generally, the nervous system. The topic of study is so complex that it requires disparate basic, clinical, and applied disciplines to cover the terrain. Within neuroscience are crosscutting paradigms—general perspectives that underlie a range of theories and methodologies in the field. The fulcrum for some of these perspectives rests squarely under constituent structures at different levels of organization, whereas the fulcrum for others emphasizes the functions of the brain and the nervous system. Illustrative of the latter are the complementing and increasingly complex fields of (a) behavioral, (b) cognitive, and (c) social neuroscience.

In (a) behavioral neuroscience, the nervous system and brain are viewed as instruments of sensation and response. Research representing this perspective tends to focus on topics such as learning, memory, motivation, homeostasis, sleep and biological rhythms, and reproduction—and on the neural mechanisms underlying these behavioral functions. (b) Cognitive neuroscience emerged as a distinct functional perspective in which the brain is viewed as an information-processing organ, with a focus on topics such as attention, perception, representations, decision making, memory systems, heuristics, reasoning, and executive functioning—and on the neural mechanisms in the human brain that underlie these representations and processes.

(c) Humans are fundamentally a social species, however, and social species, by definition, create organizations beyond individual members. These superorganismal structures evolved hand in hand with psychological, neural, hormonal, cellular, and genetic mechanisms to support them because the consequent social behaviors helped these organisms survive, reproduce, and care for offspring long enough so that they too reproduced, thereby ensuring their genetic legacy. Social neuroscience represents yet another broad perspective that seeks to understand these neural, hormonal, cellular, and genetic mechanisms and their relationship to psychological, behavioral, and social levels of organization.

For instance, according to the "social brain" hypothesis, the complexities of the social environment contributed to the complex structure and function of the primate brain. As such, the perspective of social neuroscience extends the traditional approach in the neurosciences, in which the solitary organism has been the focus of investigation, by additionally examining the social entities and systems of which the organism is a member. In order to accomplish this goal, social neuroscientists apply a multilevel approach to examine factors at levels of organization ranging from the molecular to the cultural.

In this entry, we will discuss some of the methodologies and disciplines utilized in social neuroscience,

while highlighting some of the current research in the field.

Interdisciplinary Perspectives

Given the complexity of the subject matter, social neuroscience is an inherently interdisciplinary field that emphasizes animal models as well as studies of humans. Through studies of nonhuman social animals such as honey bees, rodents, and primates, social neuroscientists have examined the biological mechanisms of social behavior through a range of manipulations, including pharmacological, neurophysiological, and genetic. Owing to the common dependence of members of social species on extraorganismal structures, the neural and neuroendocrine mechanisms that have been shaped by social demands are often conserved across species. Animal models, therefore, play an important role in the field because they permit a broader manipulation of biological factors and the collection of brain data than would be possible in normal human participants. For example, studies of the monogamous prairie vole and the highly related, but polygamous, montane vole have led to the discovery that variations in neuropeptide systems, specifically the oxytocin receptor system, mediate the variation in social structure between the species of vole, such that the higher the receptor density in reward areas in the brain of prairie voles, the more monogamous their behavior. This work has led to research indicating that polymorphisms in genes coding for the oxytocin receptor play a role in the sociality of numerous organisms, including birds, primates, and humans.

Humans are also a unique social species in that our social institutions, civilizations, and cultures are highly developed; our territorial reach knows few boundaries; and our selection of and impact on the environment in which we live constitute an active and purposeful process. Studies of humans, particularly those that permit calibration with animal models, also play an important role in the field. For instance, human studies often utilize neuroimaging techniques, which are correlational in nature, or lesions, which can lack specificity. To address these limitations, teams of investigators using both animal and human participants to test theoretical hypotheses are increasingly common in the field. Admittedly, this integration is sometimes made difficult by differences in methodology and in the levels of organization that are investigated. Moreover, a biological mechanism underlying a social behavior in an animal model may be only a part of a more complex, multifarious mechanism in humans, which means that the differences as well as the similarities between human and animal research have the potential to advance our understanding.

Research on the neurotransmitter oxytocin is illustrative. As previously noted, the role of oxytocin in social interaction was observed in a number of animal studies prior to its being investigated in human studies. In animal studies, oxytocin has been observed to play a role in a wide range of social behaviors, including pair bonding and social recognition. The research in animals provided an initial understanding of the mechanisms through which oxytocin acts in neurological systems. Human studies now indicate that oxytocin is an integral part of a large number of behaviors, including trust, perceived social threats, and reactivity to stressors. These findings, in turn, have raised new questions regarding the actual neurological mechanisms that can be addressed in animal research.

Constructs, Component Processes, and Computations

Social neuroscience is complex, in part because it entails analyses across multiple levels of organization. Such analyses bring not only opportunities but also challenges. One such challenge is that there is more than one sufficient cause for many of the phenotypes or social behaviors that are of interest. Studies that focus on single causes can appear to have large effects in controlled experimental paradigms but later can be found to have small effects in naturalistic settings or when other factors are no longer controlled. For example, in studying drug abuse, the effects of opiates on an individual are dependent upon both genetic factors, such as individual variance in the opioid receptor system, and environmental factors, such as the social context in which the drugs are administered. Both factors have been found to interact with drug abuse at the level of opioid receptors in the brain, and thus both must be considered in order to fully understand the construct of interest.

The biological mechanisms underlying many social-psychological constructs, such as empathy,

attachment, or the "self," are not instantiated in the activity of a dedicated neuron or even of isolated neural regions but rather involve multifarious circuits of neural activity. This is due, in part, to differences in what constitutes a category across levels of organization. For example, lesion studies of the concept of the self have revealed that some lesions compromise episodic aspects of the self whereas others affect semantic summaries of traits. Similarly, neuroimaging studies of self-awareness and self-referential processing reveal widely distributed regions of neural activation, including the medial prefrontal cortex, ventrolateral prefrontal cortex, dorsolateral prefrontal cortex, bilateral temporal poles, and insula. Improvements in our understanding of the neural mechanisms underlying the construct of the self will likely be advanced by a better specification of the component processes that go into this complex psychological construct. For example, even a concept as specific as self-referential processing is likely to need to be decomposed into component processes (e.g., self–other differentiation, similarity to self) and neural computations before the underlying neural mechanisms are clearly delineated.

The Prospects of Social Neuroscience

Social neuroscience is a growing field with exciting potential for research that affects all aspects of life. There are clear commonalities between social neuroscience and other fields in the neurosciences, and the former provides a unique perspective in which multilevel analyses are utilized to understand the complex relationships between social behaviors and the underlying neurobiological mechanisms. The perspective of social neuroscience, which focuses on interacting brains, provides insights not apparent from analyses of a solitary brain. For instance, the notions of determinism and monism when applied to a solitary brain lead to the conclusion that consciousness is epiphenomenal. The premise of social neuroscience is that the study of the human brain is incomplete when treated as a solitary organ. When the brain is viewed within a social context—that is, within a context of interacting brains—the material mechanism by which consciousness plays a role in subsequent brain states is revealed. From the perspective of social neuroscience, language evolved to communicate and coordinate with conspecifics rather than simply to talk to oneself. When communicating with others, our brain spontaneously attempts to predict the present and future behavior of others through certain processes, such as social cognition, mentalizing, and theory of mind—although, but perhaps better than chance, we are far from perfect mind readers. This imperfection means that when we communicate our conscious states to others, our subsequent brain states are not entirely predicted by the brain states that gave rise to our conscious expression. Specifically, our subsequent brain states are in part determined by our prior brain states and in part by the brain states that result from interactions with other brains—other brains whose influence on our subsequent brain states are themselves determined in part by the communication of aspects of our consciousness. One might counter that a person's communication is itself the result of prior and ongoing brain states, so that it is the brain states, and not aspects of consciousness, that are being communicated. The case of the display screen on a computer helps clarify why this output is instrumental, even if its influence is mediated through interactions with other humans.

Thus, because our brain underlies communication with other brains, that is, because it is social, the conscious beliefs and intentions we communicate to others have an impact on the brains of others that was not entirely predictable by their prior brain states, and their responses to us influence our subsequent brain states in ways not entirely predictable by our prior brain states. Consciousness in this social context may therefore have the potential to serve a functional role through a mechanism that adheres to the principles of monism and determinism.

John T. Cacioppo, Aaron B. Ball,
Greg J. Norman, Louise C. Hawkley,
and Gary G. Berntson

See also Biology and the Social Sciences; Cognitive Sciences; Cultural Evolution; Evolutionary Psychology; Mirror Neurons and Motor Cognition in Action Explanation; Neuroeconomics; Neuroethics; Neuroscience and Politics; Primatology and Social Science Applications; Social Cognition; Sociobiology

Further Readings

Berntson, G. G. (2006). Reasoning about brains. In J. T. Cacioppo, P. S. Visser, & C. L. Pickett (Eds.), *Social neuroscience: People thinking about thinking people* (pp. 1–11). Cambridge: MIT Press.

Cacioppo, J. T., & Decety, J. (2011). Social neuroscience: Challenges and opportunities in the study of complex behavior. *Annals of the New York Academy of Sciences, 1224,* 162–173.

Cacioppo, J. T., & Hawkley, L. C. (2009). Perceived social isolation and cognition. *Trends in Cognitive Science, 13,* 447–454.

Cacioppo, J. T., Hawkley, L. C., Crawford, L. E., Ernst, J. M., Burleson, M. H., Kowalewski, R. B., . . . Berntson, G. G. (2002). Loneliness and health: Potential mechanisms. *Psychosomatic Medicine, 64*(3), 407–417.

Cacioppo, J. T., Lorig, T. S., Nusbaum, H. C., & Berntson, G. G. (2004). Social neuroscience: Bridging social and biological systems. In C. C. Morf, A. T. Panter, & C. Sansone (Eds.), *The SAGE handbook of methods in social psychology* (pp. 383–404). Thousand Oaks, CA: Sage.

Cacioppo, J. T., & Patrick, B. (2008). *Loneliness: Human nature and the need for social connection.* New York, NY: W. W. Norton.

Norman, G. J., Cacioppo, J. T., & Berntson, G. G. (2009). Social neuroscience. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(1), 60–68.

Social Norms

The Latin word *norma* appears in slightly varying orthographic forms in many languages: in Spanish and Portuguese as *norma;* in Russian as норма; in English, Swedish, and German as *norm* or *Norm;* and in Finnish as *normi*—with the sense of a rule, standard, pattern, or model. Social norms are rules; in a wide sense, they may be defined as patterns of behavior accepted, expected, or required in a certain group. Social norms determine how a citizen or a member of a group may act or ought to act in various situations. Norms do not exist in isolation but as elements of norm systems, whose function is to guide the behavior of a population of norm-subjects.

This entry first introduces what norms and norm-contents are; then focuses specifically on social norms, reviewing various kinds of norms; and ends by explaining social norms in contradistinction to social conventions.

Norms and Norm-Contents

Conceptually, we should distinguish between a norm, the content of a norm, a norm-formulation, and the way the norm-content is presented to a norm-subject or some other audience. Norms are temporal entities that begin and cease to exist within a population of norm-subjects, but a norm-content is an abstract entity, a normative (*deontic*) proposition that can be shared by different norms belonging to different systems of norms. A norm-content is expressed in a norm-sentence (norm-formulation), and it can be presented in different ways—normatively (prescriptively), to direct the norm-subjects' behavior, or descriptively, as information about the content of a system of norms in force in a certain group.

Social Norms

Social norms are distinguished from legal and moral norms, and they must also be distinguished from practical norms, which inform an agent about the necessary or best means of achieving his or her objectives and satisfying his or her interests. Social norms are informal rules; unlike legal norms, they lack an authoritative codification and are not created by governmental acts of promulgation. In this respect, they resemble moral rules. There is a great deal of overlap among social, legal, and moral norms; both social scientists and moral philosophers are interested in the *mores* of societies, and socially significant norms tend to acquire an authoritative codification in the legal system of a society. Social scientists are interested in providing descriptive and explanatory accounts of the norms in various groups, and books on etiquette and good manners are attempts to formulate a code of behavior for the "best society." The author of such a book may acquire the status of an authority on good manners, and the book is then regarded as if it were an authoritative codification of socially acceptable or desirable forms of behavior.

Cristina Bicchieri has defined the concept of social norm in terms of two conditions, which may be called the *belief condition* and the *conditional-preference* condition. According to these conditions, a rule of behavior R is a social norm in a population or group P if and only if (1) almost all or sufficiently many members of P believe that almost all members of P comply with R and (2) almost all members of P prefer to comply with R, provided that almost all other members of P comply with R. The belief in the existence of a norm may also involve the belief that other members of P prefer one to comply with the norm and that failure to comply may engender

sanctions. This account shows two significant features of social norms. First, social norms, like many other social entities, are partly constituted by the shared beliefs and intentions of the members of a population or social group, the norm-subjects. And second, a norm can exist in a group even in the absence of general compliance if each subject expects and prefers the other members to comply with the norm and believes them to have similar beliefs. For example, there is a strict social requirement for marital fidelity among U.S. politicians, even though most of them might in fact be unfaithful to their spouses.

Kinds of Norms

Norms can be divided into mandatory norms, permissive norms, and power-conferring norms. *Mandatory norms* are requirements (obligation norms) or prohibitions; the content of such a norm is an ought-proposition, a proposition to the effect that the subject ought to or has an obligation to do or not do something. Noncompliance with a mandatory norm is associated with a ("negative") sanction or a threat of a sanction; the sanctions may be severe or not more than expressions of social disapproval, and they tend to become internalized among the norm-subjects. Mandatory norms can be said to function as exclusionary reasons, that is, reasons for disregarding other possible reasons for action in a situation to which the norm applies; thus, they help coordinate behavior in a group and increase its predictability. Social scientists often use the term *sanction* for "positive" sanctions, that is, rewards, as well as "negative" sanctions. If there is a reward attached to the performance of an action (type) A but no negative sanction attached to its omission, A cannot be said to be governed by a mandatory norm: In such a case, A is a meritorious action, not a required action. We might say that A is governed by a *merit norm* and regard merit norms as a subclass of permissive norms.

Bicchieri's characterization of social norms applies to mandatory or quasi-mandatory norms, that is, norms according to which a certain action is required by R in a situation S. *Permissive norms* may provide exceptions to mandatory norms or simply describe socially acceptable courses of action. Both permissive and mandatory norms are often associated with certain positions or roles; for example, only parents may punish their children for bad behavior. A *power-conferring norm* gives a norm-subject the normative power to create norms for other subjects. These norms are also usually attached to certain positions or roles; for example, according to some systems of social norms, people can create new rules of behavior for their own children, but they do not have the same normative power with respect to the children of others.

Social Norms and Conventions

The concept of social norm is closely related to that of social convention, and it has been argued that social conventions are norms, regularities to which the subjects who follow the convention believe they ought to conform; thus, conventions have normative force. According to David Lewis's (1969) well-known analysis of the concept of (social) convention, a convention is an equilibrium solution to a coordination problem, that is, a combination of the actions of different agents (subjects) in which each agent has done as well as he can, given the actions of the other agents. Once such an equilibrium has been established, it tends to become a norm: As Lewis argues, "One is expected to conform, and failure to conform tends to evoke unfavorable responses from others" (p. 99). On the other hand, not all norms are conventions. The existence of a convention requires general (but not necessarily universal) compliance with the convention among the subjects, but as was pointed out above, a social norm can exist without general conformity. Thus, conventions, unlike social norms, can be said to be essentially *behavior dependent*. Moreover, it is clear that not all social norms can be seen as solutions to coordination problems. For example, some dietary norms may have originated from the practical interest to avoid illness, and such practices tend to retain their normative force even when the original reason for them has ceased to exist. There are a multitude of theoretical models based on rational choice theories and evolutionary game theory for explaining how norms and norm systems emerge, are preserved, and change (see, e.g., Axelrod, 1986; Bicchieri, 1993 & 2006; Skyrms, 1996; Ullman-Margalit, 1977).

Risto Hilpinen

Author's note: I wish to thank Ms. Anneli Hilpinen, MBA, for discussion and advice concerning this entry.

See also Collective Agents; Collective Goals; Collective Values; Commitment; Common Knowledge; Conventions, Logic of; Cooperation/Coordination; Cooperation, Cultural Evolution of; Social Conventions; Social Rules

Further Readings

Axelrod, R. (1986). An evolutionary approach to norms. *American Political Science Review, 80*(4), 1095–1111.

Bicchieri, C. (1993). *Rationality and coordination.* Cambridge, England: Cambridge University Press.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms.* Cambridge, England: Cambridge University Press.

Cubitt, R. P., & Sugden, R. (2003). Common knowledge, salience and convention: A reconstruction of David Lewis's game theory. *Economics and Philosophy, 19,* 175–210.

Gilbert, M. (1989). *On social facts.* Princeton, NJ: Princeton University Press.

Gilbert, M. (2008). Social convention revisited. *Topoi, 27,* 5–16.

Hilperin, R. (2006). Norms, normative utterances, and normative propositions. *Análisis Filosófico, 26*(2), 229–241.

Horne, C. (2001). Sociological perspectives on the emergence of norms. In M. Hechter & K.-D. Opp (Eds.), *Social norms* (pp. 3–34). New York, NY: Russell Sage Foundation.

Lewis, D. (1969). *Convention: A philosophical study.* Cambridge, MA: Harvard University Press.

Posner, R. A., & Rasmusen, E. B. (1999). Creating and enforcing norms, with special reference to sanctions. *International Journal of Law and Economics, 19,* 369–382.

Post, E. (1922). *Etiquette in society, in business, in politics and at home.* New York, NY: Funk & Wagnall.

Raz, J. (1980). *Practical reason and norms.* Oxford, England: Oxford University Press.

Skyrms, B. (1996). *Evolution of social contract.* Cambridge, England: Cambridge University Press.

Southwood, N., & Eriksson, L. (2011). Norms and conventions. *Philosophical Explorations, 14,* 195–217.

Ullman-Margalit, E. (1977). *The emergence of norms.* Oxford, England: Clarendon Press.

von Wright, G. H. (1963). *Norm and action.* London, England: Routledge & Kegan Paul.

Social Objects Versus Technical Objects

The category of *objects* has received much attention in recent years. Various contributions, such as the actor-network theory of those such as Bruno Latour and the object-oriented philosophy of Graham Harman, have done much to return objects to the very center of social theorizing.

The specific focus of this entry is on the terms *technical object* and *social object*. Each of these terms has been used in various ways. For example, the term *technical object* features prominently, if quite differently, in the works of philosophers of technology, such as Martin Heidegger and Gilbert Simondon, while the term *social object* appears in a range of accounts from symbolic interactionism (referring to objects that we "give meaning to in our interactions") to marketing (where social objects are those objects that "bring us together").

The focus of this entry is specifically upon uses of the terms *technical object* and *social object* that have come to the fore in attempting to capture, or elaborate, a particular distinction, namely, that between activities such as using a hammer to mend a shutter and using a passport to gain access to another country. Although both kinds of activity involve the use of material objects for some clearly defined purpose, the first activity involves only, or primarily, the use of the physical properties of the object itself (e.g., the weight or hardness of the hammer), whereas the second activity involves the use of something other, or much more, than this.

In considering the nature of this distinction, two main questions arise. First, exactly what kinds of *causal* powers and properties are we drawing upon when objects such as hammers, nails, machinery, passports, identity cards, money, and so on, are used, and where are such powers and properties located? Second, are such objects the same kinds of things, being used in different ways, or is there a useful *distinction* to be made between the *kinds* of objects that can be captured by the terms *technical objects* and *social objects?*

This entry identifies three strands of social theory that either directly address or have important implications for the nature of the distinction noted above

and the questions raised. These strands are distinguished by a focus upon each of the following: (a) the intrinsic and extrinsic causal powers of objects, (b) the technical and social function, and (c) the signaling properties of different objects.

Intrinsic and Extrinsic Causal Powers

In the first strand, it is argued that we use objects by harnessing their causal powers or properties and that different objects can be distinguished in terms of the kinds of causal powers they possess. Thus, when we use a hammer to mend a shutter, we are primarily using powers *intrinsic* to a hammer to perform the task at hand, such as its weight, the proportion of its handle to its head, and so on. Crucially, the properties of the hammer that we are harnessing, such as the hardness of iron and steel relative to wood or plastic, are pretty much the same in all cultures and at all times; these properties do not depend upon social relations.

In contrast, when we use social objects, we appear to be drawing upon much more than the object itself. Using a passport to go through customs involves the harnessing of *different* powers from those possessed by the passport as a physical object; whole sets of relationships (between citizens and nation-states, between states, between the particular traveler and the customs official at the gate) are drawn upon when using a passport. In this sense, we can say that much of what is harnessed is *extrinsic* to the object itself; the user harnesses properties that some object has by virtue of its relation to other things, people, collectivities, and so forth.

One particularly developed example of this position is to be found in the work of Clarence Ayres. Ayres, drawing upon the work of Thorstein Veblen, is primarily concerned with the role of technology in economic progress and, in particular, with a "dichotomy" between technology and institutions. At the heart of the dichotomy is a distinction between what Ayres terms *tools* and *icons* (or sometimes *fetishes*). Whereas tools and icons are similar in that they are material "things" used to accomplish certain ends, the "effectiveness" of an icon depends upon the social status, standing, or relationships of the tool user, while a tool is an artifact that will perform in much the same way whoever uses it.

From this basic distinction, Ayres elaborates an ontology of objects, which he uses to account for differences in growth and development. Technical objects (tools) have a range of properties (combinability, travel, durability, etc.) that give them a transcultural and dynamic character. Social objects (icons and fetishes) have properties too, which can be harnessed in a variety of different ways, but the (extrinsic) causal powers of such objects are not isolatable from the communities within which they have meaning and potency. For Ayres, the use of social objects tends to put a break on economic development, where their use entrenches conservative and "past-binding" attitudes and ways of doing things. In contrast, the use of technical objects, much more conducive to recombinability and so increased technological dynamism, tends to lie at the heart of increased technological and economic development.

The focus in Ayres's account, however, is primarily upon technical objects, and the particular manner in which the extrinsic power of social objects is thought to operate is left largely unaddressed.

Technical and Social Functions

A second, quite different strand maintains that the difference between activities such as mending shutters and passing through customs is to be understood in terms of the different *uses* of such objects rather than in terms of differences in the objects themselves. More specifically, such activities can be distinguished by the kinds of *functions* (in particular technical functions and social functions) that are ascribed to the different objects in use. A well-known example of the different kinds of function ascribed is provided by Lewis Binford. The crucial point here is that the same object can always be viewed under the aspect of its technical, social, or ideational functions. However, the problem of specifying the differences between these kinds of function (under what circumstances and why one function should be understood as technical and another as social) is left unresolved.

In this context, it is worth noting the work of John Searle on social institutions. For Searle, the attempt to maintain a distinction between social and technical *objects* leads to a series of contradictions. If we admit that the £5 note is at the same time a piece of paper and legal tender, Searle argues, then we must accept that we do not have two kinds of objects but rather that there are different "facts" that we might give about a particular object.

When we use a £5 note to buy a sandwich, we are not using the powers or properties of the £5 note; rather, the power at work is what he calls "deontic power," the power of rights, obligations, and so on, which are ultimately the properties of the community within which the £5 note and the user are situated. The £5 note just happens to be something that we have given a "status" to, which allows it to *perform* a particular function. Although there may be a minimal physical requirement for an object to be given some status or function, the object has no general characteristics that might form the basis of a separate category of social objects. Rather, objects have different functions, and where the function depends upon the status we attribute to the object, we have a social function. Relations are not constitutive, and causal powers are not extrinsic (in the sense that the properties of a thing do not depend on the relations in which it stands); rather, these causal powers are external to the object and not something *possessed* by it at all.

Signaling

The third position, unlike the second, does maintain that there is a distinction to be made between social and technical *objects,* but this distinction depends upon the *signaling* capacity of the former. For example, the power of money to facilitate buying something may well be grounded in the relations of credit and debt that hold between individuals in the community, but for money to be used, it must be able to *signal* the kinds of credit relationships that exist within the community itself, and it is only on account of this ability to signal that money can be so used. Indeed, most of the examples given of social objects (passports, driving licenses, deeds, money, etc.) do seem to signal something (citizenship, the ability to drive, ownership, credit, etc.).

If it is an object's ability to signal that sets it apart as a social object, what are the implications for our attempts to locate the powers harnessed when we use such objects? Even though Searle talks of objects symbolizing or counting as other things, to talk of the power residing in the community, as he does, surely misses something. Similarly, it is not clear either that this something is well captured by the idea of extrinsic powers. In this case, we need to talk of powers that are initiated, triggered, or set in play because of the object's ability to signal, properties that depend on the kind of signaling made possible by that object. Such an effect is perhaps best captured by the idea of emergent powers that arise through signaling.

It should be clear that however intuitively plausible the distinction between social and technical objects might be, invoking the distinction raises a range of ontological issues that are both complex and central to much of the philosophy of social science.

Clive Lawson

See also Actor-Network Theory; Searle and the Construction of Social Reality; Social Construction of Reality; Social Constructivism; Social Ontology, Recent Theories of

Further Readings

Ayres, C. E. (1978). *The theory of economic progress.* Kalamazoo, MI: New Issues Press.

Binford, L. R. (1962). Archaeology as anthropology. *American Antiquity, 28*(2), 217–225.

Harman, G. (2011). *The quadruple object.* Winchester, England: Zero Books.

Heidegger, M. (1977). *The question concerning technology, and other essays.* New York, NY: Harper & Row.

Latour, B. (2005). *Reassembling the social: An introduction to actor network theory.* Oxford, England: Oxford University Press.

Lawson, C. (2009). Ayres, technology and technical objects. *Journal of Economic Issues, 43*(3), 641–659.

Simondon, G. (1989). *Du mode d'existence des objets techniques* [The mode of existence of technical objects]. Paris, France: Aubier.

Smith, B., & Searle, J. (2003). An illuminating exchange: The construction of social reality. *American Journal of Economics and Sociology, 62*(1), 285–309.

Social Ontology, Recent Theories of

Social ontology asks in what sense, if any, social entities exist and what the basic nature of social entities and social relations might be.

Recently, the focus has been on whether a social group or institution amounts to anything over and above its individual members or functionaries and their psychological attitudes and actions. This

question has been pursued in two ways. One way, which may be called *the ontology of social entities and relations,* sharpens the question of whether a group is anything over and above its individual members by using concepts of traditional and recent metaphysics to ask whether a group is identical with, reduces to, supervenes on, or is constituted by its members. The other, and currently more popular, way, which may be called *analysis of social concepts,* asks whether our everyday concept of a social group can be analyzed in terms of concepts of the individual group members and their psychological attitudes and actions. Regarding social relations, recent studies have asked whether these can be reduced to the psychological attitudes and actions of the related individuals. Can friendship, for instance, be understood as a composite of the actions, expectations, and desires of friends?

In the following sections, we will discuss these two approaches in turn.

The Ontology of Social Entities and Relations

Ontological individualism regarding groups holds that a social group amounts to nothing over and above its individual members, and group beliefs, intentions, and actions amount to nothing but admissible composites of the singular attitudes and actions of the members—the sorts of attitudes and actions recognized by belief–desire–intention psychology, for example. This thesis becomes controversial when applied to groups of the sort we would commonly call "social entities" and think of as agents—associations and social organizations—as opposed to pluralities of individuals that we do not think of in this way, like genders or ethnicities. Many individualists would allow that members of a group of the former sort act jointly, but all individualists would deny that such a group resembles a single human being in being a single subject of thought or a single agent. If groups are nothing over and above their individual members, they are not plausibly taken to achieve the unity of attitude and action required of a single agent. Although individualism is the orthodox view among analytic philosophers, it has proved challenging to formulate a defensible version of the view.

If a group is a single entity, whether understood as a single agent or not, then it cannot be identical with each of its members, since by the transitivity of identity, this would entail that each member is identical with each other member—an absurd result. This observation suggests a *plural-identity view* of a group. The suggestion is that we rely on the logic of plural terms and the notion of plural identity to identify a group with the plurality of its members. The sentence "Batman and Robin captured the Joker" does not analyze as "Batman captured the Joker, and Robin captured the Joker," since the former sentence is true and the latter sentence is false. Evidently, we should not understand the subject term "Batman and Robin" as consisting of two referring expressions, one of which, "Batman," refers to Batman and the other of which, "Robin," refers to Robin. But we may instead understand the subject term as referring plurally to Batman and Robin *as two*. A term referring to a group refers plurally to several individuals as several and not to any one of those individuals. The suggestion is then that despite appearing to be a singular subject term, "the group" does not refer to one thing but refers plurally to the members of the group. The group is plurally, not singularly, identical with these members. The plural-identity view does not by itself offer an account of the conditions under which groups remain the same in counterfactual circumstances, and until the view is supplemented with such an account, it remains unclear whether it qualifies as an individualist account treating a group as nothing over and above its members. One difficulty with the plural-identity view of a group is that distinct groups can have the same members. The members of the Library Committee and of the Food Committee of a college can be numerically identical, even though these are distinct groups. Then, identifying each group plurally with its members will violate the transitivity of identity. Plural identity might be defended here by employing the concept of relative identity rather than absolute identity: A member of the Library Committee is not identical with a member of the Food Committee relative to the sort "being a member of the Library Committee."

On the *mereological-sum view,* a group is a mereological sum of its members, where such a sum is one or another of the smallest objects that has as its parts each of the members. This view escapes the preceding difficulty with the plural-identity view if the same individuals can form distinct mereological sums. But a consequence of the view is that being a member of a group entails being a part of the group. And this leads to this implausible reasoning: My toe is a part of me; I am a member of the Chess Club

and so a part of the Chess Club; hence, by the transitivity of the "part of" relation, my toe is a part of the Chess Club—a counterintuitive conclusion. David-Hillel Ruben has given the mereological-sum view sustained critical scrutiny.

Christian List and Philip Pettit have defended the view that there are group agents quite analogous to single human agents. They focused on groups formed as a result of the prior joint intentions of individuals, such as governments and economic firms. A group counts as an agent under these conditions: Attitudes can be attributed to it under the intentional stance (i.e., its behavior is sufficiently predictable given the attribution of these attitudes); these attitudes are capable of sufficient coherence (attitude-to-fact coherence, consistency of beliefs and preferences, and attitude-to-action coherence); and there is something amiss—a fault is attributable—when these attitudes fail to be coherent. The literature on collective decision making shows that many common rules for aggregating individual beliefs and preferences, such as majority rule, lead to incoherent aggregate attitudes. Pettit and List argue that certain procedures, such as straw polling, can enforce coherent attitudes and thus provide the coherence necessary for a group agent. Furthermore, Robert A. Wilson has defended the *group mind hypothesis* in cognitive science, on which a group mind is attributed to explain behavior and cognitive capacity.

David-Hillel Ruben has argued that not all social relations can be reductively identified with nested systems of the interlocking singular beliefs, expectations, and actions of individuals. What it takes to be the mayor of a town, for example, cannot be so identified because it varies conventionally from town to town. This is a multiple-realizability objection to a property reduction, analogous to a common objection to the psychophysical identity theory in the philosophy of mind. Moreover, any specification of a candidate-nested system with which to identify being a mayor must be circular in a way that prevents a reductive identification.

Analysis of Social Concepts

Conceptual individualism is the view that talk of groups and joint activity can be analyzed as talk about individual members and their singular attitudes and actions. Seumas Miller has offered an individualist account of joint action as *interdependent interpersonal actions under a common end: A*'s action *x* and *B*'s action *y* constitute a joint action only if *x* depends on *y,* and conversely; *A* and *B* have a common end for which each performs the relevant action, and this end cannot be realized by one of the agents without the action of the other. This account leaves undefined the constitution relation between the joint action and the individual actions *x* and *y.* If that relation is simply the plural identity of the joint action with actions *x* and *y,* then it does not by itself provide the resources to explain the counterfactual conditions under which the same joint action occurs. The account would have to be supplemented with a specification of when the same joint action could be performed by more or fewer individuals than actually perform it. Without such a supplemental definition, we cannot be sure that the account remains within individualist resources.

Michael Bratman has proposed an account of shared cooperative activity, a species of joint action. Our *J*-ing (e.g., our painting) is a shared cooperative activity only if I intend that we *J* and you intend that we *J*. Bratman restricts *J* to a "cooperatively neutral" action type, excluding reference to an essentially joint action type (like playing tennis), because referring to intentions about a specific action type of this sort would employ the conceptual resources of the notion defined, shared cooperative activity, thereby introducing a circularity into the definition. This limitation on *J* raises the difficulty, like that noted for Miller's account, that the analysis does not by itself explain the counterfactual existence conditions of shared cooperative activity. Moreover, the limitation entails the risk of preventing the account from giving conditions for shared cooperative activity of all types, including essentially joint types. Under this limitation, the account will be fully general only if for every essentially joint action type, there is a cooperatively neutral action type such that satisfying the conditions is enough for a shared cooperative activity of this essentially joint type. But there is the danger that in a case like playing tennis, there is no cooperatively neutral activity *K* such that our satisfying the conditions for *K* is enough for our playing tennis. The difficulty has also been raised, by J. David Velleman among others, that my intending that we *J* must settle that we *J* for purposes of my practical reasoning, but it cannot do so because, at least when our *J*-ing is a joint action, I cannot think of myself as controlling whether we *J*.

John Searle has argued against an individualist account of collective intention. Searle allows that individuals are the only subjects having minds and the only agents, and every intentional state is a state of an individual mind. But he maintains that for a plurality of individuals to have a collective intention to perform a collective action, it is not enough for each to have a singular intention to act. Rather, each must have a "we-intention" to perform the collective action, where a we-intention is an intentional state of the individual but not a singular intention of that individual to perform any action. In addition to we-intentions of the participants, a collective intention to perform a collective action also requires that each has a singular intention to do his or her part in the collective action. Since a we-intention is an attitude of an individual, an illusory we-intention is possible: An individual can have a we-intention even though no one else has a similar we-intention and there is thus no collective intention. Margaret Gilbert has objected to Searle's account on the ground that it assigns collective intention even when coincident individual we-intentions arise from an illusion of collective intention, but such cases are not cases of collective intention. Ben and Elaine each individually thinks, "We intend to get married," and each says so to his or her parents; each assumes that the other would sincerely say this is true if questioned; thus, each has a we-intention that, on Searle's account, should together suffice for a collective intention. But if their parents discovered that the two had not even discussed the matter with one another, they would correctly conclude that there was no collective intention. Searle builds an account of status functions (e.g., a piece of paper's counting as money) and of social institutions on his account of collective intentionality by making these rest on the collective acceptance of deontic propositions (that we are enabled or required to act in certain ways).

Margaret Gilbert's early account of how a group forms supplied an item missing from Searle's account of collective intention—communication between the participants. On her account, a group forms when each of several individuals is, independently of the others, "quasi-ready" to share in an attitude or action as a body (where quasi-readiness is readiness conditional on the like readiness of others and their expression of willingness to share), and each expresses this willingness to the others. Once such an expression of willingness is secured, the individuals are jointly ready to share in the attitude or action as a body. This joint readiness is then sufficient for the existence of the group. Joint readiness entails an espousal of the goal of performing the action, and individuals act jointly just in case each acts in light of this espousal. It has been objected that the notion of sharing employed in this account of group formation is that of joint or participant action, since nothing less than a positive attitude toward such sharing would be sufficient for joint readiness; and for the same reason, the notion of action as a body must be a general notion of action that in this case is potentially instantiated by a joint action. Although this circularity parallels that of Bratman's account of shared cooperative activity, it does not similarly threaten Gilbert's account of a group, since her account does not aspire to remain within the conceptual resources of individualism.

In later work, Gilbert employs a primitive notion of joint commitment that is explained in part by elaborating on the obligations such a commitment entails. She offers an account of group formation that begins not with an individual's quasi-readiness to share in an attitude or action as a body but with an unconditional personal readiness to enter with others into a joint commitment (to hold an attitude or to act). A group forms when individuals who are personally ready in this way express this readiness with the understanding that, provided that others express their like personal readiness to enter a joint commitment, such a commitment is formed. The existence of such a joint commitment is sufficient for the existence of a group. A joint commitment to believe a proposition or intend an action is sufficient for a group belief or intention and does not require that any member of the group also believe that proposition or intend that action.

Gilbert proposes further that a joint commitment is at its core normative in entailing obligations of (and reasons for) the members to do their part in performing the relevant joint actions. A joint commitment is distinguished from a personal commitment by the fact that an individual can unilaterally form and rescind a commitment of the latter sort but not of the former sort. A joint commitment with others gives individuals a participant, as opposed to a singular, reason for following through in doing their part. Such reasons override singular reasons arising from an inclination to the contrary. Thus, for Gilbert, a key function of group membership

is to provide individuals with overriding reasons to participate in cooperative activities even when they are personally disinclined to do so. Michael Bratman has argued for the alternative view that individual obligations to participant actions arise not from joint commitments but from each participant's purposive creation in others of an expectation of the participant's participation, a creation that has value because there is value in having a certain matter settled. Caroline Baumann has argued that, contrary to Gilbert's proposal, the obligations that arise from joint commitments do not differ fundamentally from singular reasons arising from inclination: They do not derive from the will but instead create reasons for following through that weigh against those of personal inclination by raising the cost of not participating. Gilbert builds an account of political obligation on her account of joint commitment.

Raimo Tuomela has developed an account of collectivity that employs a distinction between having an attitude or acting in the we-mode and in the I-mode. The former is a matter of an attitude or action as a member of a group, the latter as a private person. Tuomela has offered an extensive taxonomy and detailed analyses of collective attitudes and actions in terms of we-attitudes and we-actions. He has given particular attention to group belief. He has also employed the concept of collective acceptance to account for social practices and institutions.

Frederick F. Schmitt

See also Collective Agents; Collective Intentionality; Collective Moral Responsibility; Group Mind; Holism, in the Social Sciences; Individualism, Methodological; Judgment Aggregation and the Discursive Dilemma; Plural Subjects; Searle and the Construction of Social Reality; Social Facts; Social Institutions; We-Mode, Tuomela's Theory of

Further Readings

Baumann, C. (2011). Gilbert's account of norm-guided behavior: A critique. In H. B. Schmid, D. Sirtes, & M. Weber (Eds.), *Collective epistemology* (pp. 227–241). Frankfurt, Germany: Ontos.

Bratman, M. (1999a). Shared cooperative activity. In E. Sosa (Ed.), *The faces of intention: Selected essays on intention and agency* (pp. 93–129). Cambridge, England: Cambridge University Press.

Bratman, M. (1999b). Shared intention and mutual obligation. In E. Sosa (Ed.), *The faces of intention: Selected essays on intention and agency* (pp. 130–141). Cambridge, England: Cambridge University Press.

Gilbert, M. (1989). *On social facts*. London, England: Routledge. (Reprinted in 1992, Princeton, NJ: Princeton University Press)

Gilbert, M. (2003). The structure of the social atom: Joint commitment as the foundation of human social behavior. In F. F. Schmitt (Ed.), *Socializing metaphysics: The nature of social reality* (pp. 39–64). Lanham, MD: Rowman & Littlefield.

List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford, England: Oxford University Press.

Miller, S. (2001). *Social action: A teleological account*. New York, NY: Cambridge University Press.

Ruben, D.-H. (1985). *The metaphysics of the social world*. London, England: Routledge.

Schmitt, F. (Ed.). (2003). *Socializing metaphysics: The nature of social reality*. Lanham, MD: Rowman & Littlefield.

Searle, J. (1990). Collective intentions and actions. In P. Cohen, J. Morgan, & M. E. Pollak (Eds.), *Intentions in communication* (pp. 401–415). Cambridge: MIT Press.

Searle, J. (1995). *The construction of social reality*. New York, NY: Free Press.

Tuomela, R. (2002). *The philosophy of social practices: A collective acceptance view*. Cambridge, England: Cambridge University Press.

Tuomela, R. (2003). The we-mode and the I-mode. In F. F. Schmitt (Ed.), *Socializing metaphysics: The nature of social reality* (pp. 93–127). Lanham, MD: Rowman & Littlefield.

Velleman, J. D. (1997). How to share an intention. *Philosophy and Phenomenological Research, 57,* 29–50.

Wilson, R. A. (2010). Social reality and institutional facts: Sociality within and without intentionality. In S. Tsohatzidis (Ed.), *Intentional acts and institutional facts: Essays on John Searle's social ontology* (pp. 139–156). Dordrecht, Netherlands: Springer.

Social Perception

Social perception, an element of social cognition, refers to the set of processes by means of which we perceive others—individuals, groups of individuals, as well as symbols—in our social world.

Social cognition, in turn, refers to the ways in which we understand our social world. Thus, social cognition refers to cognitive manipulation of information regarding conspecifics. The nature of social perception is a much debated issue, and there are views that even question its very existence, given that it is not always clearly distinguished from social cognition.

This entry looks at the nature of social perception and tries to demarcate its borderlines by focusing on its relations to social cognition and sensory perception. It ends by drawing the relation between social perception and related areas such as empathy, Simulation Theory, as well as recent uses of mirror neurons in motor- and action-cognition.

Social Perception and Social Cognition

Sensory perception, regardless of particular sensory modalities (visual, tactile, olfactory, etc.), is often understood as occurring at two stages, a passive and an active. The passive stage of visual perception, for instance, is what occurs at the peripheral parts of the human brain dedicated to perception, while the active stage involves processes of understanding what is out there. Even though no highly sophisticated cognitive processing is involved in the passive part of perception, it is widely accepted that memory, in other words, stored mental representations, does play a significant role in perception.

In an attempt to distinguish social perception from social cognition, it could be argued that a relation analogous to the one holding between passive and active perception also holds between social perception and social cognition. In this sense, social perception could be seen as part of social cognition. But that does not imply that for social perception to earn its keep, there have to be special sensory-perceptual channels dedicated to social stimuli, as, for instance, in the case of visual perception (a kind of "social sense" as it were, akin to the "moral sense" championed by the moral philosophers of the Scottish Enlightenment in the 18th century or, more recently, by contemporary philosophers like Jesse Prinz). It suffices to say that social perception occurs on the basis of a stimulus triggering our perceptual apparatuses, which then triggers deployment of a proprietary set of representations acquired during experiences with our social environment.

Clearly, in real-life situations, it is very hard to isolate social perception from social cognition, and it seems far more useful to examine the two together. But this should not be taken to mean that social perception does not exist or that it is not useful in its own right in understanding social behavioral patterns. For instance, waiting for your turn patiently instead of jumping the queue does not have to occur by virtue of sophisticated reasoning processes such as "Respecting others is generally good, therefore I should wait for my turn patiently." Most often, information acquired during social perception, for example, a set of representations of previous queuing experiences in this case, suffices in order to explain the occurrence of social behavioral patterns like the one described above.

Social Perception and Sensory Perception

In order to further understand the nature of social perception, we must have a thorough understanding of the relationship between sensory perception and cognitive processing. However, the distinction between perception and cognition is not a clear one either, and this unavoidably allows space for a "gray area" between the two. For instance, even though it is widely accepted that the way in which we get to perceive the world around us is not isolated from previous experiences, it is still intuitive to assume that there are stages of the perceptual process that are clearly noncognitive.

But is it plausible to assume that early, noncognitive, perceptual stages suffice in order to perceive social features, or is it that social perception requires more sophisticated cognitive processing?

Perception of our social world occurs in the same way as sensory, for instance, visual, perception of our physical environment occurs, namely, by virtue of light reflecting on the surfaces of objects (or other agents in the case of social perception) and exciting our sensory apparatus. During perception of a given object, a representation of that object is formed in the perceiver's mind. A plausible way to distinguish between sensory and social perception is to assume that representations of conspecifics or of social processes either bear a different weight or are tagged as different from sensory perceptual representations and are in turn stored in a distinct locus in the mind or forwarded toward distinct cognitive processes.

If this is accepted as plausible, then it could be argued that social perception differs from sensory perception. Furthermore, if it is shown that social perception occurs precognitively, then it could be said that social perception is clearly demarcated and distinguished from social cognition.

The above suggestion seems to enjoy support from independent evidence from researches on facial recognition. In particular, Edmund Rolls and colleagues showed that certain groups of cells in the inferior temporal (IT) cortex of primates are responsive to faces. Interestingly, they also found that single cells responded strongly to a few faces and showed little response to certain other faces or nonfacial stimuli. Martin Tovée and associates also found face-selective cells in the IT cortex and the cortex in the banks of the anterior part of the superior temporal sulcus of macaques. Furthermore, Edmund Rolls and colleagues found that certain neurons in the cortex in the anterior part of the superior temporal sulcus in primates responded to facial expressions and to facial movements involved in gesturing. Crucially, they also found that neurons in the temporal area were more likely to have responses related to the identity of faces. Finally, Robert Desimone and colleagues, who also studied face recognition in primates, found a population of cells in the IT cortex that responded selectively to faces, and their response patterns did not alter over changes in the stimulus's size or position in the visual field. Note that these neurons did not respond to other complex objects such as flowers and snakes.

Further evidence of the "special nature" of social perception can be found in the literature on gaze following. Among others, Rechele Brooks and colleagues found that at nine months, infants do not respond differentially to the perceptual status of the eyes. They merely follow the adult's turn of the head toward a target. In contrast, older infants sharply differentiate these two conditions and closely monitor the adult's perceptual organs.

In light of the above evidence, it could be said that there are neurons in the brain that respond selectively to "social stimuli." Note though that this does not imply that our social world is perceived, in the broad sense of the term, without deployment of cognitive processing. Rather, it suggests a way on the basis of which the raw materials of social cognition are formed. In this sense, it is plausible to assume that social perception could be seen as a part of social cognition that is distinct from sensory perception.

The Role of Social Perception

Social perception contributes greatly to our understanding of others. It does this by providing the representational inputs to processes of attributing beliefs and desires to others. In turn, this attribution process occurs by virtue of simulating the mental states of others. Simulation processes are often seen as underlain by brain areas associated with the human mirror neuron system (neuronal groups in motor cortical areas that get activated both while perceiving and while performing a given action, and are thus seen as not differentiating between specific agents). Furthermore, simulating the mental states of others involves empathizing with them. For instance, when an agent perceives a subject's facial expressions, she visually represents the expression in question, and the emotional states associated with these expressions also get activated in the beholder's mind. That is, the subject empathizes with the perceived agent by virtue of the neurons underlying the appropriate emotional state that would have caused the subject herself to draw similar facial expressions getting activated. In a sense, social perception is the starting point of all of the above processes that are involved in understanding others.

Alex Tillas

See also Embodied Cognition; Empathy; Grounded Cognition and Social Interaction; Joint Attention and Social Cognition; Mirror Neurons and Motor Cognition in Action Explanation; Simulation Theory; Social Cognition; Social Neuroscience; Theory Theory

Further Readings

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science, 8,* 535–543.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. J. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience, 4,* 2051–2062.

Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading.* New York, NY: Oxford University Press.

Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: A Hebbian perspective. *Trends in Cognitive Sciences, 8*(11), 501–507.

Prinz, J. (2007). *The emotional construction of morals*. New York, NY: Oxford University Press.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience, 2*(9), 661–670.

Rolls, E. T., Cowey, A., & Bruce, V. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas (and discussion). *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences, 335,* 11–21.

Rolls, E. T., & Tovée, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology, 73,* 713–726.

Tovee, M. J., Rolls, E. T., & Azzopardi, P. (1994). Translation invariance and the responses of neurons in the temporal visual cortical areas of primates. *Journal of Neurophysiology, 72,* 1049–1060.

Social Practices

The concept of social practices has recently become important in discussions of the nature of social life and ongoing human existence. Theories employing the concept also typically embrace an account of activity that forsakes traditional subject–object ways of thinking. Theorists of social practices have offered novel accounts of society and human existence that challenge reigning approaches, inform empirical research, and contribute to the integration of philosophy and social inquiry.

Social Practices and "Practices"

Since the 1980s, the expression *social practices* (or usually just "practices") has widely appeared in theoretical accounts of social or human life. The expression and word are sometimes used almost unreflectively to name a general type or realm of phenomena that is central to the topic under discussion. This usage signals that theorists construe their subject matters as rooted in or as forms of human activity—for common to practically all theories utilizing either term is the notion, often unarticulated, that (social) practices are bundles of actions performed by different people. A long line of key topics in the human sciences have been analyzed on the assumption that bundles of different people's actions are crucial to them. Examples are reason, mind, normativity, language, identity, science, the society–nature relationship, learning, communication, gender, organizations, consumption, and social change.

Because practices are bundles of activity, attention to them perpetuates the long-standing belief in social thought that activity is central to social life. Attention to practices is also one stream in a wider intellectual development that promotes activity as equally, or even more, central to human life as mind. Theories of social practices thereby link up with other accounts that make action central to human existence, for example, post-Husserlian phenomenology, pragmatism, and even behaviorism.

There is little to unify theories that marshal the term *practices*. Many such theories have no articulated conception of practices. Significant differences, moreover, mark different explicit conceptions: The relationship of practices to actions, on the one hand, and to social phenomena, on the other, can be variously understood. Still, conceptions of social practice exhibit common themes and have arisen on the background of particular philosophies and paradigmatic social-theoretical and philosophical accounts.

Social Life

One of these themes is the centrality of practices to both the constitution of—what they are made of—and the causality responsible for social phenomena. Theories championing this theme represent an alternative to reigning social theories that treat individuals, interactions, language, structures, systems, and so on, as the principle generic phenomenon in social life. Against, for example, those forms of individualism that build up social phenomena from the actions and mental states of individual people, theorists of social practices argue that actions inherently belong to activity bundles (practices) and only as such help constitute social entities. Almost all social theories that wield an explicit conception of practices uphold this theme, though individualist analyses of something called "social practices" also exist.

The two paradigmatic theories of social practices are those of Pierre Bourdieu and Anthony Giddens. Bourdieu conceived of society as composed of fields

of practice, where a field is a bounded domain such as agriculture, politics, recreation, or education. Practices in a field pursue the specific matters at stake in it, drawing on material, symbolic, and cultural capitals accumulated there and arising from subconscious generating mechanisms (habitus) that, in mirroring objective properties of that field, ensure that practices perpetuate those properties. Giddens, meanwhile, analyzed a slew of prominent social phenomena, including institutions, change, systems, power, and ideology, by reference to practices, which he understood as structured by sets of rules and resources.

Philosophers, too, have advocated the constitutive and causal centrality of social practices. Examples are Charles Taylor's doctrine that social reality *is* practices and Theodore Schatzki's claim that social phenomena are slices or aspects of nexuses of practices and material arrangements.

Human Activity

Theorists of social practices also usually sport a particular philosophical conception of human activity. Since the 17th century, philosophical discussions of human activity have been structured by the dichotomy between subject and object. On the background of the ideas of the celebrated 20th-century philosophers Martin Heidegger and Ludwig Wittgenstein, philosophers of a practice persuasion, such as Taylor and Hubert Dreyfus, have made two important claims. The first is that action rests on something nonpropositional, something that cannot be put into words, for example, skills or practical understanding. This nonpropositional know-how is embodied, as opposed to contained, in a subject or its mind. The second claim is that activity so understood both is conceptually prior to and underlies the traditional division between mind and the world. This claim fosters the philosophically important conception of practices as constellations of doings and the nonpropositional understandings underlying them, which form the background on which—the place where—states of mind, human activities, rules, and interpersonal relations receive determinate content—that is, are the states, activities, rules, and relations they are.

This picture of action also characterizes social-theoretical practice theories, paradigmatically, those of Bourdieu and Giddens. In Bourdieu, the nonpropositional phenomenon that underlies action is habitus: arrays of subconscious bodily structures that generate activity, thought, and perception. Meanwhile, according to Giddens, "practical consciousness"—what a person knows but cannot say—is the central agency responsible for human activity.

Theories highlighting practices share the conviction that prominent features of human or social life not previously so conceived are best understood as constituted or rooted in bundles of actions resting on embodied know-how. As the above discussion shows, the concept of social practices also joins philosophy and social theory. Practically all theories that make the concept central are resolutely multidisciplinary.

Theodore R. Schatzki

See also Embodied Cognition; Habitus; Holism, in the Social Sciences; Individualism, Methodological; Knowing-How Versus Knowing-That; Pragmatism and the Social Sciences

Further Readings

Bourdieu, P. (1990). *The logic of practice* (R. Nice, Trans.). Stanford, CA: Stanford University Press.

Dreyfus, H. L. (1991). *Being-in-the-world.* Cambridge: MIT Press.

Giddens, A. (1979). *Central problems in social theory.* Berkeley: University of California Press.

Schatzki, T. R. (2002). *The site of the social.* University Park: Pennsylvania State University Press.

Taylor, C. (1985). Interpretation and the sciences of man. In *Philosophy and the human sciences: Philosophical papers* (Vol. 2, pp. 15–58). Cambridge, England: Cambridge University Press.

Social Rules

Social rules are the rules of social groups. Different groups may have different rules. A possible social rule is the rule that one is not to talk on a cell phone while dining with friends. Although such rules are commonplace, theorists disagree on what precisely they amount to. There is pointed disagreement over the attitudes individual members of a social group must have if there is to be a social rule. One account

argues that there is a social rule when members *personally* accept a certain pattern of action as a standard for the group. Another account argues that *joint* acceptance by the members is required. The joint account claims to explain better how people respond to rule breakers.

This entry briefly reviews three prominent accounts of social rules and highlights their differences.

H. L. A. Hart on Social Rules

According to the British legal philosopher H. L. A. Hart, there is a social rule within a group if and only if, roughly, all or most group members (a) regularly conform to a particular pattern of behavior; (b) consider this pattern a standard to which group members *ought* to conform, all else being equal; (c) pressure one another to conform to the rule; and (d) think that such pressure is justified. Though influential, Hart's account is open to criticism. The necessity of each of his conditions has been questioned. Furthermore, it seems that there are situations that meet all of his conditions but do not instantiate the concept of a social rule.

Thus, consider the following case: All members of a particular group are regularly truthful, they consider *not lying* to be a standard to which group members ought to conform, they pressure one another not to lie, and they believe that such pressure is justified. As described, not lying seems to meet all of Hart's conditions for a social rule of this group. But though each group member *individually* considers not lying to be a standard to which group members ought to conform, it is not clear that it is a rule *of the group*.

It has also been argued against Hart's account that the kind of pressure put upon rule breakers, including demands for conformity and rebukes for nonconformity, requires a special standing or authority. Hart's conditions could be satisfied without group members having that authority.

David Lewis on Social Convention

Some see *social conventions* as a species of social rule. According to David Lewis, conventions are patterns of behavior conformed to by members of a given group within a recurring *coordination problem*. Here is a sample coordination problem: Sue and Tom agree to meet at "the Greek restaurant downtown." Later, both realize that there are two Greek restaurants downtown. They have no way to contact one another. Tom and Sue have a coordination problem. Each wants to go to the same restaurant as the other, and neither cares which restaurant that is, but where should each one go?

According to Lewis, a group has a convention if and only if, roughly, there is a pattern of behavior in the context of a particular coordination problem such that all or most members of the group conform to that pattern, expect one another to conform to it, and prefer to conform to it on condition that the others do, and all of this is known to all.

A clear case of a Lewisian convention is driving on the right side of the road. When all relevant persons prefer to drive on the side everyone else drives on, everyone expects everyone else to drive on the right, everyone does so drive, and all this is known to all.

Driving on the right may have become the convention by chance. Perhaps some people started driving on the right for no particular reason and others took it from there. Lewis emphasizes that there need be no explicit agreement in order to start a convention, nor need the parties be moved by a sense of their obligations to others. Group members conform to conventions given their personal preferences and their personal expectations that others will conform.

Does Lewis's account of convention capture our everyday understanding of social rules? One problem with the account is that not all social rules seem to be grounded in coordination problems. The cell phone rule imagined earlier seems to be an example. It may simply make sense to some people to have such a rule. If that is right, Lewis's account is in at least one respect too narrow to account for social rules generally.

Another problem is that Lewis's account seems unable to explain important aspects of social rules. People think of the rules of their group as something that members should conform to regardless of personal preference. Furthermore, Lewis's account seems not to entail that group members have the standing to rebuke one another for failing to conform to an established convention.

Margaret Gilbert on Social Rules

Margaret Gilbert's account of social rules differs from those of both Hart and Lewis in significant ways. It does not appeal to what individuals personally

accept, expect, or prefer from others. It invokes the *joint* acceptance of a rule and explains this in terms of something akin to an agreement.

If two individuals make an agreement, then if one violates the agreement without release from the other, that other has grounds for rebuking the violator. In other terms, agreements create obligations of the parties, one to another. When there is a social rule in Gilbert's sense, the parties are in this way obligated to conform to certain standards of behavior.

Gilbert's account of social rules is not in terms of agreements as such but rather in terms of something she takes to be the result of agreement making. It can also occur independently of the making of an agreement strictly speaking. This is what she refers to as *joint commitment*.

According to Gilbert's account, in her technical terminology, a given pattern of behavior is the rule of a particular group if and only if members *jointly commit themselves to accept as a body* that they are to conform to it. Those who make such a joint commitment are said *jointly to accept* that they are to conform to the said pattern. As a result of this process, each member of the group is committed to conform to the rule in question, and no one is in a position unilaterally to rid oneself of this commitment; the permission of the other group members is required. Furthermore, Gilbert argues, each member is in a position to rebuke other members for nonconformity to the group's rule and to demand conformity when it is threatened. Thus, she sees her account as an improvement over that of both Hart and Lewis in this respect.

Margaret Gilbert and Maura Priest

See also Collective Agents; Commitment; Conventions, Logic of; Plural Subjects; Promises and Agreements; Rule Following; Social Conventions; Social Facts; Social Norms

Further Readings

Dworkin, R. (1977). *Taking rights seriously.* London, England: Duckworth.

Gilbert, M. (1999). Social rules: Some problems for Hart's account, and an alternative proposal. *Law and Philosophy, 18,* 141–171.

Gilbert, M. (2008). Social convention revisited. *Topoi, 27,* 5–16.

Hart, H. L. A. (1961). *The concept of law.* Oxford, England: Clarendon.

Lewis, D. K. (1969). *Convention.* Cambridge, MA: Harvard University Press.

Social Studies of Science and Technology

Social studies of science and technology is a field of scholarship that has developed in its current form only in the past 40 years or so; it is often referred to as STS, for science and technology studies. In its broadest sense, STS is concerned with the conceptual and empirical analysis of science and technology in their social context.

Orthodox philosophy of science and STS are clear contenders with regard to the study of science, and to a large extent, STS has challenged received assumptions about science and technology. Equally, STS can be seen as important for the development of the philosophy of the social sciences, as certain social-scientific disciplines, notably sociology, have been expanding their erstwhile domain by looking at science, the role of scientific theories, and scientific production, while at the same time positing interesting philosophical questions about science and the way it should be studied. In this vein, the reader may also look at a parallel development, the study of the economics of scientific knowledge (see entry in this encyclopedia), as an offshoot of modern developments in the philosophy of science and of social science.

Though there are many reasons for being interested in the social context of science and technology, there are three analytical issues that are most prominent and of conceptual importance for philosophy and the social sciences.

(1) The first of these revolves around the question of the extent to which science and technology—but especially science—are bound up with or independent of their societal context. As will be explained below, this issue was the central point of contention in early scholarly debates around STS in the 1970s and into the 1980s. (2) The second analytical issue relates to the growing importance of science and technology for the social and human sciences, given the extent to which contemporary political and

policy issues themselves hinge on scientific matters. Perhaps, the key global policy issue of the 21st century is climate change, and of course, policy debates on climate change revolve precisely on scientific models and the credibility of scientific expertise. Unanticipated to an extent in the early years of STS, the social role of science and technology is now at the center of politics and socioeconomic analyses. Finally, and in some ways linking the previous two points, (3) there is a major academic and policy debate about the extent to which science and technology can (or should) be democratized and opened up to public participation.

Strong Programs and the Distinctiveness of Science

Sociological interest in science and technology can plausibly be traced back to the early 20th century and before. Certainly, public and scholarly debates in the 1930s and 1940s concentrated on the extent to which developments in science were driven, and indeed molded, by changes in society and the economy. But the work that is now taken as the core of "modern" STS was produced some three decades later. The pivotal arguments at the time were directed against prevailing assumptions in the philosophy and history of science.

The then-orthodox position was that science fundamentally progresses in a rational manner—following the precepts of the scientific method—and that the path of scientific development is significantly immune from social or cultural influence. Of course, a hostile cultural environment can slow the advance of science. On the other hand, exceptional support for research in one area (astronomy perhaps) can influence the relative development of different scientific disciplines. But scientific methods disclose the nature of the material world in objective ways, so that one would not expect sociocultural influences on the content of scientific claims themselves. Similarly, in cases of scientific disagreement—controversies have broken out frequently in the history of science—one would expect the controversy to be ultimately resolved by the production of the better evidence. In these ways, science was taken to be exceptional among humans' intellectual products. Theological study has, for example, been undertaken for millennia, and there are countless works of theological analysis. But theological knowledge is not cumulative in the way science is taken to be; on the contrary, it is assumed that cultural influences can be detected in the way theological views change. The same would be true of metaphysics or aesthetics but not of science.

Philosophers of science generally subscribed to one variant or another of this orthodox view, a view with which many practicing scientists were also content. The history of science tended to be conducted in isolation from other historical scholarship, and it charted the way in which the forefront of scientific understanding advanced through time.

Perhaps the most telling challenge to this orthodoxy came not from a social scientist or anyone committed to founding STS but from an unusual historian of science, Thomas S. Kuhn. Most celebratedly, in his 1962 book *The Structure of Scientific Revolutions,* Kuhn presented historical material arguing against key aspects of this exceptionalism (of science vis-à-vis the rest of human knowledge) on the orthodox view. He proposed that scientific thinking was characterized not by steady, cumulative advance but by the existence of *paradigms:* fundamental commitments and exemplars that framed how knowledge was made. Paradigms—such as Newtonian physics—organized what it was possible to think at any particular period. And, according to Kuhn, the move from one paradigm to another was more like a political revolution than a scientific debate. Although Kuhn famously used the term *paradigm* in a number of differing ways and although he offered no strict methodological basis for his studies, his fundamental insight took hold.

Launching Programs

In the following decade, two sets of British sociologists began setting out their standpoints. Both were Kuhn-friendly, though not precisely Kuhnian. Both developed programmatic positions. And both were more radical than Kuhn. Probably, the more well-known position is the "Strong Program" set out by David Bloor at the University of Edinburgh (also a contributor to this encyclopedia). His arguments were generally supported by other scholars who were then based in Edinburgh, for example, Barry Barnes and Steve Shapin, who were clearly within the Strong Program, even if they did not follow Bloor's preferred route in doing the sociology that followed from the Strong Program. What was

strong about Bloor's Strong Program was its insistence that sociology should treat all kinds of knowledge equally. The social scientist should adopt the same "impartial" approach to explaining people's beliefs about science or mathematics as he or she would adopt for analyzing ideas about religion or political ideology. Even more radically, this equality of treatment should be extended to the explanation of beliefs that come to be regarded as true or as false. For Bloor, the sociology of knowledge would be "symmetrical" in its style of explanation, although Bloor was clear that this meant that the same *types* of cause would explain true and false beliefs. He did not assert that it was only sociological or cultural factors that explained beliefs.

The other programmatic position was associated with Harry Collins and his then-colleagues at the University of Bath, and was known as EPOR, the Empirical Program of Relativism. Collins developed his position in relation to the analysis of contemporary, not historical, science. His fundamental argument is that if one studies any live scientific controversy or dispute, the facts of the matter are precisely what is up for grabs. Therefore, the truth of beliefs cannot be part of the explanation of the outcome, since the truth is not known by anyone until the outcome has been determined. If the scientists involved in a controversy do not know the truth (and they wouldn't be having a controversy if they did), the sociologist plainly has no way of knowing what the truth is either. By necessity, sociologists will thus have to be impartial and symmetrical. Similarly, in Collins's view, good historians studying past controversies will interpret them in terms of the knowledge available to the participants at the time and not by reference to views that subsequently come to be taken to be correct. Accordingly, the only appropriate stance for the sociologist is a relativistic one, treating the competing views as equally legitimate. However, this is only a *methodological relativism,* a position appropriate to the study of controversy. It does not imply that the sociologist is necessarily a thoroughgoing relativist.

Collins and colleagues generally chose to study controversies in the physical and experimental sciences, suggesting that if their approach was successful there it was assured of success in "softer" knowledge domains. The EPOR view was developed on the basis that studies of scientific controversy indicated that scientific data can always be interpreted in differing ways (this came to be called "interpretative flexibility"). Given that, in principle at least, this interpretative flexibility would allow disputes to carry on indefinitely, EPOR proposed that controversies are in practice resolved through social processes (referred to as processes of "closure") and not through evidential necessity. Roughly speaking, closure occurs when people choose not to carry on arguing or they run out of the resources needed to do so; this point was similar to Kuhn's views on the succession of paradigms. Finally, it may sometimes be the case that there is a connection between these processes of social closure and social forces beyond the immediate community of scientists; closure may, for example, be related to a broader struggle for power or credibility, though the process is always social, even when not societal in this broader sense.

Both of these positions, the Strong Program and EPOR, joined Kuhn in making the very content of scientific knowledge the subject of social-scientific study. Both, thereby, lent credence to a kind of constructionist view in the analysis of science. Scientific beliefs emerged historically not as the gradual disclosing of the underlying makeup of the natural world but as versions of nature assembled out of cultural resources.

This broad constructionist position was firmly in place by the 1980s, even if some of the conceptual issues around the meaning of relativism or about Bloor's "causes" of belief were unresolved. Several contemporary authors were critical of the relativist or perceived antirationalist approach implied; they argued that the programmatic Edinburgh and Bath views were not as compelling as their proponents suggested. Others tried to offer compromise positions—including a proposed "weak program."

Similar arguments were applied to the field of technology studies. In the 1980s, Trevor Pinch, a colleague of Collins, applied the EPOR approach directly to the analysis of technology, swapping EPOR for SCOT (the Social Construction of Technology). As with EPOR, he argued, when there are disputes over rival technologies, it is not simply the case that the objectively superior technology wins out. Rather, "closure" around a technology, while apparently driven by technical considerations, is a social process. This argument was in some ways regarded as less radical than EPOR since it is hard for anyone to maintain that technologies are free of

societal influences. All technologies are, by definition, constructed, if only in an everyday sense, so a constructionist position on technology is less counterintuitive than one on science.

After, or in Place of, Constructionism

In practice, many analysts in social studies of science and technology have settled for less strident versions of constructionism than those apparently offered by Bloor and Collins. Such authors are somewhat relaxed about the symmetry and impartiality demands and are content to argue that there are elements of construction in the bits of science they are concerned with, without necessarily maintaining that scientific beliefs are just as constructed as beliefs in other fields. For example, detailed studies have lately been completed of current disputes over "infanticide" among primates. The idea here is that in the wild, newly dominant male primates may kill the offspring of former rivals. This is a controversial idea since it suggests that humans' near-relatives may be routinely murderous. But the controversy persists—pretty much as EPOR proposes—because such killings are hard to observe and because those episodes that stand the best chance of being observed relate to primate populations that are living in the least natural circumstances. In this case, there are compelling reasons for impartiality and symmetry between the two possibilities (that infanticide is "normal" or that infanticide occurs only by accident or because of abnormal circumstances), but there is no need for analysts of this case to advance general claims about relativism in relation to the rest of science.

Mild constructionist views have also been adopted in practice by many adherents of actor-network theory (ANT), an approach devised in slightly differing ways by Bruno Latour and Michel Callon in Paris. In its programmatic versions (popularized in the 1980s), ANT made its own radical turn. It introduced a further kind of symmetry, arguing that there is no inevitable validity in distinguishing between human and natural actors (an asymmetry that is central to both EPOR and the Strong Program). On this ANT view, controversies are to be understood as conflicts between competing heterogeneous networks—that is, networks indiscriminately made up of "social" and "natural" elements. In Latour's celebrated study of Louis Pasteur and anthrax, this meant that Pasteur's views won out because Pasteur created successful alliances with vets and some farmers, and with livestock and vaccines. He defeated human opponents and microbes. ANT very cleverly draws attention to the way in which scientific claims are enmeshed in institutions and relationships. But in the hands of lesser exponents, the ANT approach amounts to little more than the retelling of customary stories about the history of science, decorated with some apparently radical terminology. The fundamental issue of what accounts for the strength of strong alliances is seldom confronted.

A further strand of work in social studies of science and technology, also distinct from the constructionist approaches, comprised studies of scientific knowledge from the feminist and political economy standpoints. Such approaches were typically in agreement with the overall notion that scientific knowledge was constructed, but they were opposed to the requirements of symmetry and impartiality. However, their opposition to impartiality was not the usual, realist one (that scientific knowledge is manifestly correct, so that there are no grounds for treating claims symmetrically) but an ethical or political judgment. Scientific claims were of most interest to such authors insofar as such claims tended to legitimate social inequalities (scientific assertions about the supposed naturalness of, say, male infidelity or of female lack of competitiveness). And it was this diagnosis of the legitimatory power of science to which they could not be impartial.

By the early 1990s, this loose federation of science studiers had become well institutionalized in many U.S., Canadian, and European universities, and particularly in North America, the overarching claims about the constructedness of scientific knowledge came to the attention of some high-profile, established natural scientists. Dismayed at what was apparently being taught about the nature of science, under their very noses on campus, these natural scientists launched a counterblast that initiated what came to be known as the "Science Wars." Despite some tactical successes (including a notorious episode in which a spoof science-studies paper, admitted by its [physicist] author—Alan Sokal—to be nonsensical, was accepted for publication in a peer-reviewed cultural studies journal), the Science Wars came to no definitive resolution, though the conflict may have made STS scholars more careful about their academic practices and ambitiously general

claims. Since the 1990s, detailed and academically robust work on the sociology of science and technology has continued to thrive.

Social Studies of Science and Technology in the "Knowledge Society"

As mentioned above, there is a kind of irony in the fact that despite sociologists having been very influential in the post-Kuhnian version of STS, their concerns were not those of the sociological mainstream at the time. Bloor and Collins, for example, had their most famous arguments with philosophers of science and rationalist historians. And the subsequent Science Wars also drew attention away from societal issues to questions about the truth and rationality of science. Yet, at the same time and to an extent apparently unanticipated by nearly all the leading figures in STS, it was becoming clear that many of these concerns about scientific rationality or about the resolution of scientific controversies were being played out at a practical level in emerging policy issues.

A series of major, yet contested, policy issues arose from around this time that very frequently looked like public manifestations of the very issues that sociologists and philosophers had been fighting over. There was of course nuclear power and efforts to demonstrate its safety (or dangerousness) in objective terms. In the United States, in particular, more and more complex calculations of risk were used to regulate policies for exposure to potentially harmful substances, from industrial solvents to electromagnetic radiation. There were lengthy international arguments over the scientific causes of "acid rain" and a parallel set of negotiations over ozone depletion. These controversies intensified in the 1990s with contests over "mad cow" disease and genetically modified crops and food, and with the gradual march of climate change to the center of global public policy. As suggested above, these debates represented an adapted and public form of scientific argumentation. Sociologists of science began to see that the strategies they had devised for analyzing disputes over 18th-century optics or farm animal anthrax in the Victorian era applied also to controversies over the ways in which genetically modified crops might or might not be a source of environmental or dietary harms.

Sociologists have always been attracted to the aspiration of typifying the age in which they live. And several sociological labels for society at the close of the 20th century picked up on society's scientific and technological aspects. People wrote widely of the knowledge society, and as early as 1986, Ulrich Beck (a contributor to this *Encyclopedia*) had characterized modern society as a "risk society." Anthony Giddens and Beck both developed this idea into a focus on society's increasingly reflexive character, with—for example—older questions about whether something was safe being replaced with arguments about how safety is proven. In this context, social studies of science and technology were clearly in the sociological mainstream in an unprecedented way, even if some STS authors were not.

As indicated above, there are almost countless examples that demonstrate how STS has impinged on current policy issues, but perhaps one stands out. This stems from a legal case that arose late in the 1980s, when a suit was filed against the pharmaceutical company Merrell Dow by two children and their parents, alleging that the children's birth defects had been caused by the fact that, during pregnancy, the mothers had taken a prescription antinausea drug (Bendectin) manufactured by the company. The company produced scientific experts who argued (essentially on epidemiological grounds) that there was no rigorous evidence that the harms were caused by the drug; it was unfortunate, they said, but such birth defects do occur occasionally. The families wanted to introduce different scientific experts, bringing alternative forms of evidence, for example, about the pathways through which such a drug would be likely to cause developmental abnormalities. The judge had to decide whether to admit the families' experts; they were excluded. The case ended in the 1990s, after being appealed all the way to the U.S. Supreme Court, which was asked to determine not the facts of the case but the correct basis for admitting scientists into court. The Court was essentially faced with a classic question in STS: What are the identifying characteristics of science?

A great deal of advice was offered to the Court, including some by STS scholars, though in the end, the Court appeared to take the most heed of the affidavits of leading scientific institutions, coming up with four indicators of scientific standing. These indicators included the idea that scientific views will typically be testable and will have been subjected to peer review and publication. Anyone familiar with the STS literature will have worked out that these

indicators are unlikely to provide much assistance. In scientific controversies—even in the public disputes over genetically modified organisms—both sides typically favor testing, yet they disagree about what makes a good test. And, as anticipated, what has happened is that the four indicators—now more commonly known as "criteria"—have themselves become the focus of a whole new paralegal industry advising on how to make one's evidence fit the criteria for legal admissibility.

In this example of reflexive modernity (the application of law to science), it turns out that STS has particularly acute insights to offer.

Public Participation in Science and Technology

STS has brought important new insights to one further public policy issue, the issue of the extent to which science and technology can or should be democratized. Since the middle of the 20th century, there has been a major presumption in favor of democratization wherever possible in Western, liberal societies. But certain areas have been exempted from this trend, notably those where scientific (or similar) expertise is taken to be crucial. Loosely expressed, nearly everyone agrees that democratic methods are appropriate for selecting the government, but more or less, no one thinks that these methods should be used to vote on the value of the gravitational constant. But the late 20th century (and the start of the present century) witnessed numerous cases that fall uncomfortably between these paradigmatic instances.

There are those examples where the scientific community wishes to do something—such as to create "admixed embryos" for medical and biological research—where the ethical desirability of doing so is contested. These cases are complicated and important in a substantive sense, though they are not particularly intricate at an epistemological level, since they tend to leave the technical aspects of scientists' expertise unquestioned and focus instead exclusively on the legitimacy of their ethical reasoning.

More philosophically complex are those cases where there is uncertainty or disagreement about how to arrive at legitimate, publicly important knowledge. Under such circumstances, neither paradigmatic possibility may apply. For example, it may turn out that the presumed scientific experts are not as well informed as they believe themselves to be. Numerous studies in STS have documented cases where the established scientific authorities have had limited insights and where, by contrast, local people with experiential knowledge or activists with a commitment to rival approaches have in some senses known better. On the other hand, there are numerous cases—particularly in the biomedical field—where nonscientific citizens have claimed a right to speak, not out of general arguments about democratic entitlement but because they have privileged access to some phenomenon, as sufferers from a chronic condition or as full-time carers. In practice, one can say, experts are less expert than they assume, and certain ordinary people are often more expert than is presumed.

Such cases are commonly presented as focusing on opening up science to more public participation, but this "opening up" is not primarily on democratic grounds but, rather, because of the acuity of the understandings that lay actors may have. STS has provided a rich analytical vocabulary for making sense of this move toward participation in science and technology. It has contributed to the major academic and policy debate about the extent to which science and technology can (and should) be democratized and opened up to public engagement and has provided practical insights into the methods that can appropriately be used.

Steven Yearley

See also Actor-Network Theory; Economics of Scientific Knowledge; Evidence-Based Policy; Knowledge Society; Philosophy of Expertise; Relativism in Scientific Theories; Science and Ideology; Social Constructivism; Sociology of Knowledge and Science; Strong Program in the Sociology of Scientific Knowledge; Technological Convergence; Technoscience and Society; Value Neutrality in Science

Further Readings

Bijker, W. (1995). *Of bicycles, bakelites and bulbs: Toward a theory of sociotechnical change*. Cambridge: MIT Press.

Bijker, W., Bal, R., & Hendriks, R. (2009). *The paradox of scientific authority: The role of scientific advice in democracies*. Cambridge: MIT Press.

Bijker, W., Hughes, T., & Pinch, T. (Eds.). (1987). *The social construction of technological systems: New directions in the sociology and history of technology*. Cambridge: MIT Press.

Bloor, D. (1991). *Knowledge and social imagery* (2nd ed.). Chicago, IL: University of Chicago Press. (Original work published 1976, London, England: Routledge & Kegan Paul)

Collins, H. M. (1992). *Changing order: Replication and induction in scientific practice* (2nd ed.). Chicago, IL: University of Chicago Press (Original work published 1985, London, England: Sage)

Epstein, S. (1995). The construction of lay expertise: AIDS activism and the forging of credibility in the reform of clinical trails. *Science, Technology and Human Values, 15,* 495–504.

Hacking, I. (1999). *The social construction of what?* Cambridge, MA: Harvard University Press.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago, IL: University of Chicago Press. (Original work published 1962)

Latour, B. (1988). *The pasteurization of France.* Cambridge, MA: Harvard University Press.

Latour, B., &Woolgar, S. (1986). *Laboratory life: The construction of scientific facts.* Princeton, NJ: Princeton University Press. (Original work published 1979)

MacKenzie, D. (2011). The credit crisis as a problem in the sociology of knowledge. *American Journal of Sociology, 116,* 1778–1841.

Rees, A. (2009). *The infanticide controversy: Primatology and the art of field science.* Chicago, IL: University of Chicago Press.

Shapin, S. (1995). Here and everywhere: Sociology of scientific knowledge. *Annual Review of Sociology, 21,* 289–321.

Yearley, S. (2005). *Making sense of science: Understanding the social study of science.* London, England: Sage.

Yearley, S. (2006). Bridging the science-policy divide in urban air-quality management: Evaluating ways to make models more robust through public engagement. *Environment and Planning C: Government and Policy, 24,* 701–714.

SOCIOBIOLOGY

Sociobiology is the study of social behavior, including human social behavior, from an evolutionary perspective. It is grounded in the theory of evolution presented by Charles Darwin (1809–1882) in his *Origin of Species,* published in 1859, but it came to fruition only 100 years later, in the 1960s. It proved to be highly controversial, particularly as applied to humankind, but it survived the onslaughts and today is a flourishing part of the evolutionary spectrum.

Background

In his *Origin of Species,* Darwin presented his theory of evolution through natural selection. First, relying on the ideas of the political economist Thomas Robert Malthus (1766–1834), Darwin argued that there is a universal population pressure, leading to an ongoing struggle for existence or, more particularly, to a struggle for reproduction. Darwin argued further that this struggle leads to an ongoing winnowing of organisms, or natural form of selection, and given enough time, this eventuates in permanent organic change. Most important from the Darwinian perspective is the fact that organisms develop characteristics, known as *adaptations,* that help them in the struggle. Paradigmatic examples are the hand and the eye.

Darwin recognized from the first that behavior is part of the evolutionary scenario. It is no good being big and strong if you do not have the inclination to fight as needed or perhaps to flee in the face of predators. Of particular interest to Darwin was social behavior as evinced in social insects like the bee and the ant. It is well known that these organisms have sterile castes of workers. How could natural selection, which is brought on by the struggle for reproduction, produce animals that do not reproduce at all? Eventually, Darwin argued that at some level, it is the nest or family that counts. Sterile workers are rather like parts of the whole rather than entities entirely unto themselves.

In the *Origin,* Darwin said virtually nothing about humankind, but some 12 years later, in the *Descent of Man* (1871), Darwin turned to our own species. He made it clear that he thought that we are the product of selection—although, increasingly, he relied on a secondary form of selection, sexual selection (involving the struggle for mates)—and he included here our social behavior and inclinations, including our moral inclinations; our religious yearnings; our aggression; as well as our sexual behaviors and dimorphisms.

For a number of reasons, Darwin's thinking about the evolution of social behavior went essentially undiscussed and unexplored for 100 years. First, there was little interest in natural selection generally, and it was not until the coming of Mendelian

genetics (developed by Gregor Mendel (1822–1884), the Austrian scientist and friar) at the beginning of the 20th century that evolutionists finally could see how selection is the overwhelming force of evolutionary change. Second, behavior generally raises a number of particular problems, not the least of which is the problem of observation and experiment. It is much easier to study the morphology of an already dead organism than the activities of organisms in nature, and it is notorious that organisms in captivity often do not behave as they do in the wild. Third, the rise of the social sciences and their somewhat jealous attempt to claim the study of behavior entirely for themselves meant that evolutionary approaches were viewed unfavorably and dismissed.

Emergence of the Field

Nevertheless, even by the 1930s, especially on the continent, things were beginning to change. The so-called ethologists started to take an interest in the evolution of social behavior. However, their work was sadly marred by a naive understanding of the working of natural selection. The reason why Darwin agonized over the social insects was because he could see that selection will almost always favor the individual over the group. This is not ideological but simply that selfishness pays over being helpful (or what biologists call being "altruistic"). The selfish individual exploits the altruist. Thus, it does better in the struggle; and so then the altruist goes extinct. Darwin, as we saw, got around this problem by considering the group of relatives—the family—as one unit. The ethologists ignored these worries and so plunged into "group-selective" explanations rather than "individual-selective" ones, thus vitiating much of their work.

It was not until the 1960s that the importance of an individual-selectionist approach was fully appreciated, and as soon as this happened, especially given that by now the general theory of selection leading to adaptation was on an increasingly firm theoretical and empirical foundation, things took off in a major way. Most significant were a number of new models showing how natural selection can promote social behavior. First, the then-graduate English student William Hamilton developed his idea of *kin selection*. Relatives share the same genes, and thus inasmuch as a relative reproduces, one reproduces oneself. Hence, help given to relatives—in other words, altruism—can in fact be a form of evolutionary enlightened self-interest and promoted by natural selection. Second, the then-graduate American student Robert Trivers developed his idea of *reciprocal altruism*. Basically, you scratch my back, and I will scratch yours. Help given to others can be of value if it ensures that at times of need others will help you.

With these models in place, as well as (thanks particularly to the English evolutionist John Maynard Smith) a growing recognition that game theory (developed in the Cold War to anticipate the moves of the enemy) has much to say in the study of animal social behavior, the way was now opened for empirical workers to search nature and start experimenting to see if indeed social behavior can be brought about by natural selection working at the individual level. There were a number of outstanding achievements in this direction, notably the studies of Geoffrey Parker on the mating behaviors of dung flies and the studies of Tim Clutton Brock and his associates on the social behavior of the red deer on islands off the coast of Scotland.

By the mid-1970s, it was apparent that finally social behavior had been brought fully within the Darwinian paradigm. As if to celebrate this fact, the subject was given its own distinctive name of *sociobiology*. Two outstanding works of synthesis and explanation appeared at this time. First, there was the magisterial overview of the subject, *Evolution: The New Synthesis,* by the Harvard entomologist Edward O. Wilson. He discussed in some detail the new models of understanding and the need for them and then gave a magnificent overview of the new subject, starting with the slime molds and then moving up through the insect world to the reptiles and then on to the mammals and birds. (Birds were becoming a particular subject of interest, given both their ease of study and the recognition that the dangers of breeding in trees meant that the males necessarily made a much greater contribution to social life than do mammals.) Second, there was the outstanding popularization of the subject, *The Selfish Gene,* by the Oxford-trained behavioral biologist Richard Dawkins. He proved particularly skilled at explaining complex notions, for instance, those drawn from game theory, without the need of heavy-duty mathematics.

At first, work in the blossoming field of sociobiology did not focus on humankind, but it was not long before students of the subject were

extending their vision to our species. The final chapter of Wilson's *Sociobiology,* "Man: From Sociobiology to Sociology," dealt explicitly with *Homo sapiens*. Wilson argued that human social structure is, for instance, a reflection of the fact that males tend to be more aggressive and females more domestic, more "coy." And this is a direct function of Darwinian selection. Then, three years later, Wilson devoted a whole volume to our species. *On Human Nature* was a full exposition of the worth of using natural selection to explore topics such as religion, ethics, and aggression. Thus, for instance, Wilson argued that religion has a biological function, namely, that it promotes group cohesiveness. Likewise, following Darwin himself in the *Descent,* Wilson argued that morality has no further basis than cementing the social bonds between fellow members of the human species.

Dawkins also extended sociobiology to humankind, although, unlike Wilson and others, he did not remain exclusively at the biological level. Rather, he suggested that there are units of culture, which he called *memes,* analogous to the units of heredity, genes. He argued that we have a form of Darwinian evolution operating pretty much exclusively in the cultural realm, somehow above and sitting on Darwinian evolution in the biological realm, and that this explains human nature and belief. Unlike Wilson, who although not personally religious appreciates its social value, Dawkins denies that religion has any social or cultural value. He explains it in terms of memes that hop from human to human, parasitically as it were. He has indeed likened religion to a disease, a form of unfortunate cultural virus.

Criticism

It was not long before critics started to appear. Expectedly, some of the objections came from social scientists. Sociobiology, especially as applied to humans, was threatening to their own autonomy and activities. Perhaps somewhat less expectedly, some of the objections came from the sociobiologists' fellow biologists—in Wilson's case, even from fellow members of the same biology department at Harvard. These biologist critics had a number of motivations, but two of the most prominent, Richard Lewontin and Stephen Jay Gould, were open in their Marxist affiliations. They clearly thought that the biological approach to humankind was threatening to their philosophy.

The criticisms generally fell into one of two camps. First, there were the epistemological objections. These included worries that there is insufficient evidence for much that is claimed by sociobiologists, especially in the human realm. Connected to this objection—perhaps not entirely consistently—was the charge that, in any case, sociobiology is unfalsifiable. Notoriously, Gould argued that sociobiologists simply provide stories akin to those fabulous tales invented by the English writer Rudyard Kipling. Gould opined that much of sociobiological thinking failed to rise above the level of "Just So" stories. You think of an idea and simply make up the evidence for it.

Also at the epistemological level were the twin objections that sociobiology is unduly reductionistic (in the sense of focusing on the part and not the whole) and deterministic (in the sense of seeing all as predetermined). Too much emphasis is put upon the gene and not enough on the whole organism, and organisms—particularly humans—are being portrayed as marionettes on the end of strings controlled by genes. (Actually, the favorite image was of a gray-suited, White businessman, on strings manipulated by the double helix.) Most particularly, the implication of human sociobiology supposedly is that we can never escape our genetic destiny. Things like IQ are fixed, and no amount of social engineering is going to change them.

The second kind of criticism was more social or moral. It was argued that human sociobiology is a thinly veiled excuse for the worst excesses of capitalism; it is grossly sexist inasmuch as it gives females very different and inferior roles to males; and worst of all, it is grotesquely racist. Gould, particularly, was eloquent on this last charge (a charge incidentally closely connected to the epistemological charge of determinism). In his book *The Mismeasure of Man,* Gould claimed to show that human sociobiology is but the end point of a dreadful tradition of putting down people who are not pure Anglo-Saxons. He disliked, particularly, the ways in which biology had been used to denigrate Jews.

Counters

Naturally enough, the sociobiologists gave back as good as they got. Studies in the nonhuman world grew apace, backed by increasingly sophisticated theoretical models. A good example, combining

theory and observation, was work done on "local mate competition," another idea from the ever-fertile mind of William Hamilton. If the offspring of one sex (probably males) compete for mates, then from the parental perspective, one needs fewer of that sex—after all, when it comes to grandchildren, one son is generally as good as another—and one should skew the sex ratio in favor of the other sex. Extensive work on the Panamanian fig wasp, where males emerge before females and brothers do gather and compete for the chance to reproduce, shows that Hamilton's prediction proves true. There are fewer males than expected and more females. Multiply this many times—and sociobiologists would argue that they have done just this in the past 30 years—and the critics' claims about "Just So" stories simply are not true.

Nor are the claims about falsifiability. Hamilton himself thought that one can explain the sociality of bees and ants by virtue of the fact that here one finds that females have both mothers and fathers, whereas males are born from unfertilized eggs and thus have only mothers. This means that sisters are more closely related (75%) than mothers and daughters (50%), and Hamilton thus thought that kin selection suggests that worker females are better off sterile and raising fertile sisters than fertile themselves and raising daughters. Now, however, we know that this is generally not true—queens are often multiply fertilized, and this destroys the expected relatedness ratios. Other explanations have had to be sought to explain sterile workers. (A favorite one is—given the dangers from without—in terms of the advantages of working harmoniously with relatives.). So falsifying is possible.

There is much work also on humans. One of the most rightly celebrated studies, by Wilson's student Sarah Hrdy, looks at family relationships. There is a well-established theorem (chiefly attributed to Robert Trivers) that suggests that high-status females will tend to have sons whereas low-status females will have daughters. The reasoning behind this is that females in nature almost invariably are going to get fertilized, whereas males will have to compete with other males. Females will therefore tend to have some limited guaranteed offspring, whereas males may have many offspring but also may have few or none. Hence, if you can guarantee your son a good start in life, you might skew toward sons, but if not, you will skew toward daughters. Hrdy shows, in some detail, how this theorem explains some otherwise puzzling behaviors in the Indian caste system. High-status families tend to have sons (presumably because of infanticide of the females), whereas low-status families tend to have daughters—or at least, they care for the daughters and leave the sons to fend for themselves.

What about reduction and determinism? Someone like Dawkins, who focuses on the gene, is a reductionist and is proud of it. On the other hand, he and his fellow sociobiologists would argue that this is hardly something that occurs at the expense of the whole. Work like Hrdy's may start at the level of the gene, but it rapidly expands to cover the whole of society. Moreover, Dawkins and his fellows would warn that uncritical enthusiasm for the whole over the part can lead one badly astray. Group selection in some sense clearly favors the group or whole over the individual or part, but generally it is not the right explanatory strategy. As far as determinism is concerned, one must tease apart the various meanings of this term. If one means some generally metaphysical thesis about the whole of nature being determined, then this may be true, but it is accepted by the Marxist as much as by the sociobiologist. If one means that in some extraspecific sense humans are determined, then it is by no means clear that sociobiology implies that this must be so. We may perhaps be biologically determined in something like our sexual orientation (this is controversial, but many sociobiologists would accept that this is so), but it does not follow that we are determined always to act on our inclinations. Sometimes people decide to have sex, and sometimes they don't. That is our individual choice.

What of the moral issues? No one can deny that sometimes some pretty hurtful and hateful things have been said in the name of biology. Darwin's views on the sexes in the *Descent* make for embarrassing reading in our age. But again, the response is that this is not necessarily always the case when one uses biology to explore human nature. Hrdy, for instance, rather reverses the role of the sexes in her analysis of human beings. In a book provocatively and informatively titled *The Woman Who Never Evolved,* she argues that the concealed ovulation of the human female means that males can never be sure that they have fertilized their mates and that therefore there is a biological reason for them to stay around, both protecting their investment as it were and then raising the offspring. If anything, it is men who are being manipulated, not females. Analogous arguments can

be brought to bear on other cases. Referring again to sexual orientation, the fear might be that if one finds gay genes, this will open the way to selective abortion of future folk with a same-sex orientation. On the other hand, equally, one might suggest that since the orientation is God-given—more precisely, evolution-given—it is hardly a matter of shame or moral condemnation and that one should simply recognize it for what it is, namely a nonmoral variant like hair or eye color (and probably a lot less easy to change or likely to infect others).

Conclusion

At the animal level today, everyone accepts that sociobiology is a straightforward and thriving part of the evolutionary spectrum, along with other areas such as biogeography, paleontology, and systematics. At the human level, undoubtedly, there is still controversy, so much so that practitioners now tend to conceal their activities under other more neutral names like "evolutionary psychology" and "human behavioral ecology." Perhaps, given human nature, this kind of tension will long endure. The only firm forecast one can make is that, like it or not, the Darwinian evolutionary approach to humankind is here to stay. The aim now must be to improve rather than refute.

Michael Ruse

See also Biology and the Social Sciences; Cooperation, Cultural Evolution of; Cultural Evolution; Determinism; Eugenics, Old and Neoliberal Theories of; Evolutionary Ethics; Evolutionary Game Theory and Sociality; Evolutionary Political Science; Evolutionary Psychology; Genetic Indeterminism of Social Action; Human Cultural Niche Construction and the Social Sciences; Primatology and Social Science Applications; Reduction and the Unity of Science

Further Readings

Buller, D. J. (2005). *Adapting minds: Evolutionary psychology and the persistent quest for human nature.* Cambridge: MIT Press.

Dawkins, R. (1976). *The selfish gene.* Oxford, England: Oxford University Press.

Gould, S. J. (1981). *The mismeasure of man.* New York, NY: W. W. Norton.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London Series B: Biological Sciences, 205,* 581–598.

Hamilton, W. D. (2001). *Narrow roads of gene land: The collected papers of W. D. Hamilton.* Oxford, England: Oxford University Press.

Hrdy, S. B. (1981). *The woman that never evolved.* Cambridge MA: Harvard University Press.

Hrdy, S. B. (1999). *Mother Nature: A history of mothers, infants, and natural selection.* New York, NY: Pantheon.

Kaldis, B. (2011). Nature and society. In G. Delanty & S. Turner (Eds.), *Routledge international handbook of contemporary social and political theory* (pp. 312–327). London, England: Routledge.

Levins, R., & Lewontin, R. C. (1985). *The dialectical biologist.* Cambridge, MA: Harvard University Press.

Lumsden, C. J., & Wilson, E. O. (1981). *Genes, mind, and culture: The coevolutionary process.* Cambridge, MA: Harvard University Press.

Orzack, S. H., & Sober, E. (Eds.). (2001). *Adaptationism and optimality.* Cambridge, England: Cambridge University Press.

Rosenberg, A. (1981). *Sociobiology and the preemption of social science.* Oxford, England: Blackwell.

Ruse, M. (1979). *Sociobiology: Sense or nonsense.* Boston, MA: Reidel.

Ruse, M. (2003). *Darwin and design: Does evolution have a purpose?* Cambridge, MA: Harvard University Press.

Ruse, M. (Ed.). (2009). *Philosophy after Darwin: Classic and contemporary readings.* Princeton, NJ: Princeton University Press.

Segerstråle, U. (2000). *Defenders of the truth: The sociobiology debate.* Oxford, England: Oxford University Press.

Trivers, R. (2002). *Natural selection and social theory: Selected papers of R. Trivers.* Oxford, England: Oxford University Press.

Wilson, E. O. (1975). *Sociobiology: The new synthesis.* Cambridge, MA: Harvard University Press.

Wilson, E. O. (1978). *On human nature.* Cambridge, MA: Harvard University Press.

SOCIOLINGUISTICS

Sociolinguistics is the branch of linguistics that explores the *social significance of language variation.* It examines the influence of membership in social categories, such as age, gender, social class, and ethnicity, on the use of language. One of its achievements has been to show that much of the linguistic variation that is found in all communities is

not random but systematic. Although no two individuals are identical, we share some characteristics with other family members, with people of our own age and sex, with members of our surrounding community, and so on. For many people, the way we dress, the food we eat, the ways in which we amuse ourselves are likely to be fairly consistent with how others in our community behave. In the same way, how we speak is affected by our place in society, and this often makes it possible for a stranger to draw some inferences about our social background simply from hearing us speak. The study of linguistic differences related to social factors is the province of sociolinguistics. This entry first introduces the use of quantitative methods and goes on to present the main domains of sociolinguistics.

Quantitative Methods in Sociolinguistics

In 1963, William Labov published in the journal *Word* an article titled "The Social Motivation of a Sound Change." Labov showed that on the island of Martha's Vineyard, off the Massachusetts coast, there was a change taking place in the way in which the inhabitants pronounced the vowels in words such as *tide* and *house*. Labov found differences in the extent to which different categories of the speakers he interviewed participated in the change. Factors such as the age of the speakers, where they lived, and which ethnic group they belonged to affected their participation in the change, though attitude toward life on the island turned out to be the factor that had the strongest influence.

Labov was able to show this because of the way in which he quantified his results. He developed a scale that assigned a numerical value to each example of the relevant sounds in all the words on the tape that included these sounds. He then obtained a score for each speaker. He could then calculate an average score for a social category—for example, those of a certain age or those who lived in a certain community or were of a certain ancestry. These group scores could then be compared with those from another category. This quantitative approach allowed comparison on a range of social factors and provided a model for future investigations of linguistic variation.

Labov followed up on his own example by conducting a large-scale investigation in New York City. This became the model for other investigations of urban speech in Detroit (Michigan), Anniston (Alabama), Reading (England), Montreal (Canada), Glasgow (Scotland), Sydney (Australia), and Bahia Blanca (Argentina). These studies revealed the systematic nature of linguistic differences related to age, gender, social stratification, and ethnicity.

Age Differences

Age differences in language use have been reported in a number of sociolinguistic studies. Young people are often in the vanguard of linguistic change, though some of their innovations are not adopted by adults. Ethnographic studies of adolescent speech behavior have shown girls making greater use of language differences to establish their identity in school-based groups or street gangs. Membership in these groups is influenced by social class categories, but participation in certain group activities has more impact on girls than on boys.

Gender Differences

Early sociolinguistic discussion of gender differences occurred in the context of investigations that focused mainly on social class differences. In a number of studies, it appeared that women used more of the standard variants than men, and different explanations were offered for this. One was that women were more concerned with respectability, while men, particularly lower-class men, were more anxious to assert their toughness, perhaps as a way of distancing themselves from a feminine world. Later studies have shown that women are often the leaders in linguistic changes, thus contradicting the notion that women are generally more conservative in language usage. In one study, the females used many more pronouns than the males, in particular the pronoun *she*. This is probably related to the fact that the women talked more about people, while the men made many more references to places.

Social Class Differences

Social class has been shown to be related to differences in the pronunciation of vowels in a number of studies in New York City, Reading (England), Glasgow (Scotland), and Londonderry (Northern Ireland) and to differences in the pronunciation of consonants in studies in New York City, Detroit, Reading, Glasgow, Cardiff, Newcastle, and

Londonderry. Most of the early work investigated differences in pronunciation, but more recent works have investigated other aspects of language. A study in Glasgow found that working-class speakers used fewer adverbs (e.g., *definitely, seriously,* and *very*) than middle-class speakers. The middle-class speakers used more examples of passive voice and more evaluative adjectives, as well as the expressions *sort of, kind of,* and *quite.*

Ethnic Differences

Much work has been done on what is now called African American Vernacular English (AAVE). The early work on AAVE was to some extent motivated by a desire to show that vernacular forms of speech are as regular and systematic as the standard language. William Labov investigated the speech of African American males aged 8 to 19 years who participated in the street culture of New York City, and he was able to show that grammatical patterns not found in the standard language were used consistently and systematically by these speakers. Despite this work, it has been very difficult to persuade many members of the public (including teachers) that these apparently ungrammatical forms are not simply mistakes.

Studies of Latino English have shown that many of its features are not typical of the forms used by native speakers of Spanish who are learning English, so Latino English is not simply a version of a foreign accent. There are Puerto Rican, Cuban, and Dominican forms of Latino English, and they all have distinct characteristics.

Ronald Macaulay

See also Discourse Analysis; Language and Society; Semantics and Pragmatics

Further Readings

Chambers, J. K., Trudgill, P., & Schilling-Estes, N. (Eds.). (2002). *The handbook of language variation and change.* Malden, MA: Blackwell.

Macaulay, R. K. S. (2009). *Quantitative methods in sociolinguistics.* Basingstoke, England: Palgrave-Macmillan.

Trudgill, P. (2000). *Sociolinguistic: An introduction to language and society* (4th ed.). New York, NY: Penguin.

Sociology of Knowledge and Science

This entry gives an account of the domain of the sociology of knowledge and the special subdomain of the sociology of scientific knowledge, clarifies the precise kind of sociological approach involved, explains what "evidence" means in this area, distinguishes the sociological from the philosophical questions, and finally pinpoints misconceptions about the task and goals of the sociology of science. A number of concrete examples are offered.

Introduction

The sociology of knowledge is a discipline devoted to investigating the role played by conventions, institutions, traditions, and interests in the creation and evaluation of knowledge, including scientific knowledge.

Knowledge is a complex, real-world phenomenon, and it can only be addressed as an object of study with the cooperation of specialists from a range of empirically grounded disciplines, such as biologists, physiologists, psychologists, sociologists, historians, and anthropologists. No single approach can offer a sufficient account of knowledge, but all the contributions can be seen as necessary components of an adequate, "naturalistic" picture of knowledge—that is, a picture that securely grounds knowledge within space and time and sees its origin and development as a causal process on a par with any other natural phenomenon.

The Task of the Sociologist of Science

To identify the kind of contribution made by *sociologists of knowledge* when looking at, say, the achievements of scientists, it will be useful to start with a special case and then show how it can be generalized. Suppose that scientists have developed two rival theories, T1 and T2, which cover some range of phenomena, for example, rival theories in genetics or in fluid dynamics. As a result of their investigations, suppose that the scientists have accumulated a considerable body of evidence but the evidence proves to be equally balanced and supports both theories to an equal degree. A number of different things could now happen. The community could split into

two subgroups G1 and G2, with the members of G1 expressing a preference for T1 and the members of G2 for T2. The scientists in these respective groups could continue to develop and apply their preferred theory and evince confidence in its ultimate triumph. Or the scientists could maintain their unity and attempt to pursue the study of both theories while waiting patiently for any evidence that might help decide between them.

These possibilities bring to the surface a number of questions. If the original group splits, why do some scientists prefer T1 and some T2? Can these preferences be explained in terms of the properties of the two groups? For example, the sociologist may discover that most of the members of G1 are British while most of the members of G2 are German or that the G1s are physicists from elite universities while the G2s are engineers from provincial technical colleges. Further analysis of the reasoning of the scientists concerned might reveal how the intellectual traditions of these institutions find expression in divergent technical preferences and decisions. Whatever the outcome of such investigations, it will be apparent that, in the circumstances described above, the original group will have to reach some accommodation with the situation. Whether its members divide up or stay united, that accommodation will involve some form of coordinated behavior and will necessarily involve some manner of social organization. It is the sociologist's job to find out what happens and why. What contingencies of power, status, influence, solidarity, and social organization determine the accommodation that is arrived at?

It is now time to acknowledge the limitations of the example. In the hypothetical scenario described above, it might appear that the experimental evidence for T1 and T2 "underdetermines" the choices that the scientists make and this is why it is necessary to bring in "social factors." The expression *underdetermination by the evidence* suggests that "evidence" is one (perhaps partial) determinant while "social factors" constitute a second (perhaps partial) determinant. This understanding has informed many of the critical responses to the sociology of knowledge offered by philosophers. Philosophers have enjoined sociologists to "disentangle" the rational from the social and to confine themselves to those cases where the evidence and reason are insufficient to determine scientific judgment. For example, sociologists are encouraged to study cases where "social factors" are said to override "cognitive factors," as in the notorious examples of the ideological distortion of science in totalitarian states. The philosopher is thus proposing a division of labor: The philosopher will address "rational causes," and the sociologist will address "social causes." There are, however, good reasons for believing that there can be *no* such division of labor, or *none* that makes sense within the naturalistic framework identified at the outset. To establish this conclusion, it is necessary to relax the idealizations in the example and focus attention on a category that has so far been taken for granted, namely, that of *evidence*. It is necessary to press the sociological analysis into the realm of the undeniably rational in order to break down the idea that rational and social "factors" are qualitatively different kinds of thing operating in competition with one another.

"Evidence"

Sociologists and historians have examined the role of *evidence* in scientific reasoning with great care, and one result stands out clearly. No theory is ever confronted with the totality of putative evidence that is available. Evidence is always used *selectively*. Some evidence is deemed better than other evidence, some evidence is counted as more important than other evidence, some evidence is described as misleading or as coming from unreliable sources, and some claimants to the role of evidence are dismissed as mere artifact or error. And *all* evidence involves some element of *theoretical interpretation*. In other words, evidence always operates by virtue of the exercise of judgment, and the judgment is constitutive of what counts as evidence.

Just as sociologists ask why the members of one group prefer one theory to another (when the evidence is balanced), they can now ask what influences the judgments that go into the selection, interpretation, and evaluation of evidence, that is, into the *construction of the evidence itself*. Here is an example. In the 1910s and 1920s, British physicists working in aerodynamics were worried about the implications of a theorem about the behavior of an ideal fluid. Lord Kelvin had proven mathematically that a pattern of flow called the *circulation* could neither be created nor destroyed. On that basis, the British experts rejected an account of the lift of an aircraft

wing that depended on the wing creating a circulation. German engineers knew all about Kelvin's result but were unworried by it and continued to develop the circulation theory with considerable success. For them, the result was not evidence that their approach would fail. Why did the two groups react so differently? The sociologist will explore the possibility that the divergent evaluations derived from the different disciplinary traditions involved. Of course, the evaluation of a piece of evidence may involve the use of further evidence, but this only raises the same questions again. It delays but cannot evade the need to take judgments into account and to explain their covariation with social categories such as nationality and academic discipline.

Misconceptions About the Sociology of Science

Unfortunately, it has become routine for certain scholars to talk of sociological "critics" of science, as if sociological enquiry was an exercise in denigration. In reality, the attempt to answer sociological questions does *not* amount to an attack on science, nor does it show that evidence is necessarily distorted by the exercise of judgment. The rooted conviction that sociologists of knowledge are expressing hostility to science derives from a failure to understand the naturalistic perspective. From a naturalistic and scientific perspective, all knowledge claims are part of the causal nexus. To locate the causes of a belief, including, of course, sociological causes, does not permit the inference that the belief is compromised. If causation implied error, then all knowledge would be defective. Such considerations also show that the timeworn claim that the sociology of knowledge is self-refuting is wrong. It rests on the false premise that causation implies error.

To drive these points home, it is worth recalling the outcome of Rudolf Carnap's attempt to develop an account of the process of scientific confirmation. He wanted to know how much support a given piece of evidence provided for a given theory, and he wanted to analyze the relation in purely logical terms. Here, we can see the philosophical ideal of *disentangling* the social from the rational in action. Carnap would surely be able to give a nonsociological analysis of the nature of inductive evidence. To develop his argument, Carnap postulated a simplified model "world" with a limited number of properties and a simplified formal "language" containing a limited number of predicates. After a strenuous mathematical analysis, he was forced to conclude that there was no unique definition or measure of the degree of confirmation but an infinite continuum of confirmation functions, so that some element of *choice* was unavoidable. Furthermore, to arrive at a determinate measure of confirmation, it was also necessary to assign a value to an intriguing parameter that Carnap called λ. It transpired that λ symbolized an inductive strategy and corresponded to the willingness to take risks when making generalizations. Here, then, was the inescapable element of *judgment*.

What might be understood as an attempt to remove human agency from the process of scientific reasoning finished up as a demonstration of its irreducible presence. Carnap himself identified λ as a psychological variable, but it is not difficult to see that in the history of science, inductive strategies are properties of *groups* of persons and are maintained, monitored, and sanctioned at the level of the group, not at the level of the individual. Inductive strategies are *institutions* sustained by the *group*. Now the sociologist can complete the process of bringing the analysis back to earth by asking, in real cases, how a group implicitly or explicitly arrives at its risk-taking strategy.

Concluding Remarks

The aim in this entry has been to describe the general principles behind the sociology of knowledge rather than the results of particular empirical case studies, though it must be emphasized that these studies represent the true center of gravity of the field. Many of them concern episodes in the history of science—for example, early scientific disputes about the interpretation of experimental findings derived from the air pump, disputes over the proper scope of genetics, or disputes in the history of aerodynamics. Their practical value is like the practical value of all studies in history, in whatever field: They increase self-awareness. Contemporary judgment can be both challenged and refined by appeal to analogies between past episodes and present problems.

Clearly, philosophy, which is not an empirical discipline, stands in a problematic relationship to the sociology of knowledge. The idea of a division of labor between the study of rational and social causes has been discussed above and rejected.

In confronting the problem of finding a role for philosophers, it is often argued that philosophers are concerned with what ought to be believed rather than what is actually believed; in other words, philosophers' concerns are exclusively evaluative. From the naturalistic standpoint, however, such a position is untenable. Why are the evaluations made in one way rather than another? Could it be that they are simply rationalizations of current practice and therefore, as justifications, wholly circular? If not, then what is the source of their credibility and authority? It would seem that from a naturalistic standpoint, philosophers can neither work alongside sociologists, dividing different aspects of knowledge between them, nor transcend the sociology of knowledge.

David Bloor

See also Objectivity; Observation and Theory-Ladenness; Science and Ideology; Social Constructivism; Social Studies of Science and Technology; Strong Program in the Sociology of Scientific Knowledge

Further Readings

Barnes, B., Bloor, D., & Henry, J. (1996). *Scientific knowledge: A sociological analysis*. Chicago, IL: University of Chicago Press.

Bloor, D. (2011). *The enigma of the aerofoil: Rival theories in aerodynamics, 1909–1930*. Chicago, IL: University of Chicago Press.

Brown, J. (Ed.). (1984). *Scientific rationality: The sociological turn*. Dordrecht, Netherlands: Reidel.

Carnap, R. (1949). *The continuum of inductive methods*. Chicago, IL: University of Chicago Press.

Harwood, J. (1993). *Styles of scientific thought: The German genetics community 1900–1933*. Chicago, IL: University of Chicago Press.

Shapin, S., & Schaffer, S. (1985). *Leviathan and the air pump: Hobbes, Boyle, and the experimental life*. Chicago, IL: University of Chicago Press.

SOLIDARITY

The word *solidarity* derives from Roman law, where *obligatio in solidum* referred to the group liability of joint debtors. This is the sense of the French word *solidarité* in the *Encyclopedia* of 1765 and in Napoleon's Civil Code of 1802. Around the 1840s, the term was adopted in German and English, was politicized, was adapted to the social sciences, and came to be used in the broader sense of emotionally and normatively motivated readiness for mutual support, as in the slogan "One for all and all for one."

Given its widely different meanings, the concept has been used in four main contexts. First, it is employed in the context of explaining or understanding the nature of social cohesion, social order, "groupness," or the "glue" that keeps societies or groups together (in the sense of social solidarity and group solidarity). Second, it has been linked to the ideal of *fraternité,* as a desirable feature of societies, political communities, or welfare states (in the sense of social solidarity and civic solidarity). Third, it denotes an attitude or demand relevant in struggles for liberation and against injustice or oppression (in the sense of political solidarity, workers' solidarity, Black solidarity, and women's solidarity). Fourth, it is employed as a universalistic ethical ideal of responsiveness to the human moral standing (in the sense of human solidarity, moral solidarity, and global solidarity).

This entry focuses especially on the first use—more relevant to the theme of this encyclopedia—and then briefly discusses the other three.

Group Solidarity and Social Solidarity

In sociology and social psychology, solidarity has been conceived either as an irreducible macro-level phenomenon of group cohesion or order or as a micro-level behavior, namely, in the form of emotions and attitudes explaining such a cohesion. The macro-level cohesion or unity can be sustained, for example, by coercion, by self-interest, or, as in the case of social solidarity, by a commitment to shared norms and valued social bonds. While it is irreducible to self-interest, the degree to which the social norms and institutions are seen to benefit oneself (and one's kin) and the degree to which one's own fate depends on that of the whole group may, nonetheless, partly explain the strength of the commitment to the social norms and institutions.

The term was introduced to sociology by August Comte, but the classic treatment is Émile Durkheim's distinction between the "mechanical" solidarity of traditional communities and the "organic" solidarity of modern societies. Mechanical solidarity is based on the similarity of the members and the dominance of collective consciousness over individuality.

Organic solidarity is based on the interdependence of different individuals and on the social division of labor. The Durkheimian distinction makes it possible both to acknowledge that traditional social ties are eroding (while not fully disappearing) as an effect of industrialization, urbanization, individualization, or democratization and to see a different basis for social life emerging, consistent with these processes, leaving room for individual differences.

As a micro-level phenomenon, solidarity has been conceptualized as prosocial behavior across different situations: helping and supporting in situations of need, doing one's share in situations of cooperation, showing fairness in situations of distribution of goods, avoiding breach in situations of trust, and facilitating moral repair when violations have taken place. The sociologist Siegwart Lindenberg sought to explain what makes people act in such a solidarity-based manner, as opposed to hedonistic or gain-seeking ways, in certain situations but not in others. In addition to the general formation of character or behavioral dispositions, situational cues have been shown to make a difference to behavior by affecting the salience of the solidarity frame, as opposed to those of immediate gratification or long-term gain. Solidarity is thus precarious and needs to be supported by factors that increase the salience of the solidarity frame.

For methodological reasons, such focus on behavior, excluding emotional or attitudinal elements, is adopted in empirical research (e.g., by Michael Hechter and Siegwart Lindenberg). Most theorists argue that behavior is not enough, as acting out of solidarity requires the presence of attitudes or emotions such as a sense of belonging, concern for the others' well-being, commitment to shared norms, valuing the social bonds in question, or identification with the group. As a distinct motivational pattern, solidarity combines elements of (extended) egoism and (restricted) altruism. It can be seen as a form of "we-thinking" based on collective intentionality, whose nature has been studied in recent social philosophy.

Civic Solidarity, Political Solidarity, and Human Solidarity

Solidarity has been taken to be one important evaluative feature of good societies (complementing social justice, democracy, and autonomy). "Solidarity" has made its way into the European Union Constitution and has been promoted by rival movements such as Marxism, Social Democracy, French Solidarism, Liberalism, Roman Catholicism, and Neo-Fascism. The classical Liberal and Catholic approaches stress interpersonal responsibility and solidarity as a private or personal virtue, while Marxists and Social Democrats typically stress structural obstacles, institutional solutions, and shared responsibility, so that solidarity can equally be a virtue of institutions. A moderate view is that institutional arrangements can promote and realize genuine group solidarity (e.g., via progressive taxation and social services), but if the institutional arrangements are obeyed for solely coercive or self-interested reasons, they fall short of genuine solidarity. In social policy research, the European welfare states have been seen as realizing, via institutional means, relatively high degrees of social (or "civic") solidarity and distributive social justice.

A third type of solidarity has been at stake, for example, in the international workers' movement, the Polish *Solidarność*, and various new social movements. In political solidarity, activists and members of the movement join to oppose the injustice or oppression of some group and possibly to seek support from outside (out-group solidarity). Fourth, the term is sometimes used very broadly to refer to the basic ethical concern for others, as in the work of Richard Rorty. In this context, it is often called *moral* or *human* solidarity. While solidarity is typically seen as a positive quality, various forms of solidarity have been criticized for their tendency to lead to the exclusion of outsiders, perhaps represented as enemies, or for the internal repression of individuality, autonomy, or personal responsibility.

Arto Laitinen

See also Collective Intentionality; Durkheim's Philosophy of Social Science; Social Capital; Social Institutions; Social Norms; Trust, Social

Further Readings

Baldwin, P. (1990). *The politics of social solidarity: Class bases of the European welfare state, 1875–1975*. Cambridge, England: Cambridge University Press.

Bayerz, K. (Ed.). (1999). *Solidarity*. Dordrecht, Netherlands: Kluwer.

Doreian, P., & Fararo, T. (Eds.). (1998). *The problem of solidarity: Theories and models*. Amsterdam, Netherlands: Gordon & Breach.

Durkheim, É. (1947). *The division of labor in society.* New York, NY: Free Press. (Original work published 1893)
Fetchenhauer, D., Flache, A., Buunk, A. P., & Lindenberg, S. (Eds.). (2006). *Solidarity and prosocial behavior: An integration of sociological and psychological perspectives.* New York, NY: Springer.
Hechter, M. (1987). *Principles of group solidarity.* Berkeley: University of California Press.
Rorty, R. (1989). *Contingency, irony and solidarity.* Cambridge, England: Cambridge University Press.
Scholz, S. (2008). *Political solidarity.* University Park: Pennsylvania State University Press.

Space, Philosophical Theories of

From antiquity through the Scientific Revolution, a major controversy was whether extension can exist without body. Although this is sometimes still debated, the original *Problemstellung* has been eclipsed in three respects: first, by the emergence of the field concept in physics; second, by the assimilation of space into space-time; and third, by the geometrization of gravity in the general theory of relativity. This entry traces the history of concepts of space and the philosophical debates concerning the nature of space.

The beginnings of the concept of space in Greek thought are not clear, since the term *space* is of Latin derivation and has no direct Greek origin. Plato's notion of the receptacle in his dialogue *Timaeus* played too much the role of Aristotle's later concept of prime matter to be considered to refer to space itself. Aristotle went to great lengths to develop a notion of place (*topos,* or τόπος) but never wove these places together in order to arrive at a doctrine of space. Earlier, the Greek atomists Democritus and Leucippus, in response to the single, changeless being of Parmenides, had introduced the notion of the void (*kenon,* or κευόυ) in which atoms can move. By the time of Epicurus, a global notion of space clearly emerged, according to which *kenon* refers to unoccupied locations, *topos* to a part of space occupied by a body, and *chôra* (χώρα) to a region of space through which a body is moving. This became standard in later atomism.

The word *diastêma* (διάστημα), which refers to (a finite) extension, led to the division of Greek and Latin philosophers into two camps. On the one hand were those, including Plato, Aristotle, and the Peripatetics, who rejected the possibility of *diastêma* without body, and on the other hand, were the atomists and the Stoics, who allowed for unoccupied extension. For the former camp, the universe is a finite material plenum. For the latter, the universe is infinite in extent. (The atomists believed the number of atoms to be infinite as well, while the Stoics took the world to be a finite plenum embedded in an infinite void.)

Descendants of these views emerged in the 17th century. René Descartes identified matter with extension, and hence by definition, there can be no empty space. The pivotal innovation was that Descartes permits the parts of space to move among themselves. Indeed, all physical phenomena are to be traced to the motions of this material plenum. Stationary space, in which this motion takes place, is only a mode of thought.

Following Descartes in the 17th century, Pierre Gassendi and Walter Charleton revived a Christianized version of the Epicurean worldview. To address the Aristotelian objections, they held space to be neither a substance nor an attribute but rather something with its own nature of being that arises of necessity from the existence of God, who is literally omnipresent. Isaac Newton took over this concept, calling it *absolute space.* Famously, he argued in the "Scholium" to the definitions of his *Principia* that the definition of the state of true motion or rest of a body requires the existence of absolute space. He also likened absolute space to the sensorium of God. This drew the ire of Gottfried Leibniz, who like Descartes, was a material plenist and believed in a transcendent God. For Leibniz, a body's true motion is internal to it and cannot be known (although there cannot be true motion without relative motion between bodies). Space is something ideal and relative, arrived at by hypostatizing the locations of bodies at relative rest.

The success of Newton's mechanics led to the widespread acceptance of absolute space, although it had its frequent critics. In the 18th century, Immanuel Kant initially thought that the existence of incongruent counterparts (left hands and right hands) supported Newton over Leibniz, only to argue later that space is the a priori form of outer perception. But Kant was unable to go on to develop

a natural philosophy of motion capable of bearing fruit for physical science.

Changes in physical theory led to new conceptions in the 19th century. It was primarily due to Thomas Young's and Augustin-Jean Fresnel's resuscitation of the wave theory of light in the early 19th century that the world became a plenum again, with an omnipresent optical ether conceived as stationary en masse in absolute space. The Scottish physicist James C. Maxwell, by identifying light with electromagnetic radiation, converted this into an electromagnetic ether, which in the hands of the Dutch physicist H. A. Lorentz at the end of the 19th century became dematerialized to the extent that one has an everywhere stationary ether, at each point of which exists an electric and a magnetic state. As such, it is difficult to discern whether this is a material plenum or rather Newton's absolute space endowed with further physical qualities. Max Abraham is representative in suggesting that the difference is but a choice of one's *façon de parler.*

Special relativity brought further changes by wedding space with time as a consequence of the relativity of distant simultaneity. Events simultaneous for one observer are not simultaneous for another observer moving relatively to the first. Thereby, spatial extension also becomes relative. The more natural physical ontology becomes one of space-time events rather than of spatially extended substances enduring over time. Yet further change appeared with general relativity. For one, the metric of spatial hypersurfaces of space-time no longer needs to be Euclidean but may have variable curvature, either positive or negative. For another, the metric tensor also serves as the gravitational potential, making it unclear whether space-time is an arena in which physical events occur or whether it is a kind of four-dimensional ether, as Albert Einstein sometimes described it. Other fields exist in space-time, and the metric field can exchange energy with them. However, gravitational energy cannot be localized as an energy density, as can the energy of other fields. By Einstein's famous formula for the equivalence of mass and energy, this entails that the gravitational metric can possess mass, only one cannot say how this mass is distributed. The best one can say is that the nature of space in general relativity is unprecedented. The introduction of quantum gravity promises to bring yet further modifications by giving space a granular structure at the extreme micro level of Planck length, 10^{-33} cm.

Robert Rynasiewicz

See also Metaphysics and Science; Space, Social Theories of; Time, Philosophical Theories of

Further Readings

Alexander, H. G. (Ed.). (1956). *The Leibniz-Clarke correspondence.* Manchester, England: Manchester University Press.

Algra, K. (1995). *Concepts of space in Greek thought.* Leiden, Netherlands: E. J. Brill.

Charleton, W. (1966). *Physiologia Epicuro-Gassendo-Charltoniana.* New York, NY: Johnson Reprint. (Original work published 1654, London, England: Tho. Newcomb)

Descartes, R. (1983). *Principles of philosophy* (V. Miller & R. Miller, Trans.). Dordrecht, Netherlands: D. Reidel. (Original work published 1644 in Latin as *Principia Philosophiae,* Amsterdam, Netherlands: Elsevier. Reprinted in *Oeuvres de Descartes,* Vol. 8, 1905, Paris, France: Léopold Cerf)

Earman, J. (1989). *World enough and space-time: Absolute versus relational theories of space and time.* Cambridge: MIT Press.

Einstein, A., Lorentz, H. A., Minkowski, H., & Weyl, H. (1952). *The principle of relativity: A collection of original memoirs on the special and general.* New York, NY: Dover.

Grant, E. (1981). *Much ado about nothing: Theories of space and vacuum from the Middle Ages to the Scientific Revolution.* Cambridge, England: Cambridge University Press.

Huggett, N. (Ed.). (1999). *Space from Zeno to Einstein.* Cambridge: MIT Press.

Newton, I. (1962). *De Gravitatione.* In A. R. Hall & M. B. Hall (Eds.), *Unpublished scientific papers of Isaac Newton* (pp. 89–156). Cambridge, England: Cambridge University Press.

Newton, I. (1999). *The principia: Mathematical principles of natural philosophy* (I. B. Cohen & A. Whitman, Trans.). Berkeley: University of California Press. (Original work published 1687 in Latin as *Philosophiae Naturalis Principia Mathematica,* London, England: Joseph Streater)

Penrose, R. (1991). The mass of the classical vacuum. In S. Saunders & H. Brown (Eds.), *The philosophy of the vacuum* (pp. 21–26). Oxford, England: Clarendon Press.

Sklar, L. (1974). *Space, time, and spacetime.* Berkeley: University of California Press.

Space, Social Theories of

This entry provides an introduction to a number of key social theorists who address the *production of space* and its implications for the analysis of social relations. It will thus show the importance of a renewed, radical reconceptualization of "space" in social science.

Space as a directly theorized concept has been, until relatively recently, a neglected and underaddressed aspect of social theory. The reprioritization of space as a fundamental element in the organization, structure, maintenance, regulation, and experience of social relations and of social life has generated interest within a number of social science disciplines. Space is no longer assumed to be a mere backdrop to social life, as something merely there, as natural, a void, a container, a vacuum waiting to be filled by objects or social activity. Instead, space is increasingly analyzed and understood as a product of social relations and as such is a contingent and complex process that not only reflects the needs, priorities, ideologies, aims, and practices of power per se (whether individual, institutional, or structural) but is also potentially in conflict with everyday uses, activities, and practices. Space as a product of social relations then operates in all spheres and at different scales, from the personal and micro to the macro and global levels.

Classical Social Theory and Space

The origin of much of social and especially sociological theory resides in the works of Karl Marx, Max Weber, and Émile Durkheim. Despite their ontological and epistemological differences, a recurrent theme in all their works was the attempt to make sense of the new social and material landscapes of modernity and industrial capitalism. However, the consideration of space as a fundamental factor in their analysis of the transition from feudalism to capitalism and its establishment as the dominant mode of production are implicitly, rather than explicitly, addressed in their works.

Weber, in his emphasis on the development of a rational organization of production, market, state, and bureaucracy, observes that all presuppose spatial organization. Similarly, the separation of work/household as a precondition for the emergence of capitalism and its extension to the separation of the public and private spheres have an implied spatiality. However, similar to his investigation of the role of the medieval city as a combination of fortress and market, a precursor for modern rational and urban industrial capitalism, the role of space is not developed.

Durkheim, in his treatment of the transition from feudalism to capitalism, focuses on the change from mechanical to organic social solidarity and emphasizes changes in the *spatial* organization as well as social division of society. Durkheim uses the spatial metaphors of material and moral density to explain the increasing distancing and isolation (resulting in egoism and anomie) of individuals who are cut adrift from traditional kinship ties and networks in increasingly concentrated urban centers. However, space is not explicitly theorized by Durkheim.

It is possible to identify in Marx's voluminous writing on capitalism an implicit, if not explicit, acknowledgment of space and spatial relations as a feature of the specificity of its operation and alienating experience. This includes space as a force of production, the destruction of previous spaces associated with the feudal mode of production and the creation of new spaces of capitalist production and circulation, the spatiality of the detailed division of labor (in society, in industry, as well as in the reproduction of labor, especially the spatial form of the industrial city, leading to what Engels describes as socio-spatial segregation), and capitalism as a world system.

For all these founders of social science, there is an *implicit* acknowledgment of the role of space, in their analyses of various factors, features, processes, and effects of the transition to capitalism, as the dominant mode of production but one that is not developed explicitly or in detail.

Simmel and the Sociology of Space

Georg Simmel (1858–1918) may be said to be the first sociologist of space in that in much of his work, not only in sociology and urban analysis but also in philosophy, literature, art, aesthetics, social psychology, and cultural analysis, there is a clear focus on space as a fundamental feature of the analysis of modernity through formal sociology. His essay "The Sociology

of Space" is an early contribution to the social theory of space, which, in keeping with his conception of society as reciprocal interaction, presents his five "aspects of space" (exclusivity, boundaries, fixity, mobility, and proximity and distance) as a means for investigating the significance of space for social forms. Simmel acknowledges that space shapes, and reciprocally is shaped by, the forms of social interactions that occur within it. This is a dynamic symbiotic relationship between social construction and environmental—that is, geographical—determinism, which has significance for subjective experience as well as the structural and spatial organization of the city as a whole or specific spatial forms within it, including the streets, squares, buildings, and places and spaces of consumption, leisure and recreation, or escape. Space thus features prominently and explicitly in Simmel's work on forms of social interactions as networks that constitute society.

Henri Lefebvre and the Production of Space

Henri Lefebvre (1901–1991) is considered the most influential theorist in the recent "spatial turn" in the interdisciplinary studies of space. His corpus of work on various aspects of modernity, particularly that relating to the urban, asserts the need to understand space both as a product (a thing) and as a determinant (a process) of social relations and actions. Space, for Lefebvre, is both a means and a mode through which and in which social relations and actions occur. Space therefore cannot be separated from social relations and is the product of ideological, economic, and political forces (the domain of power) that seek to delimit, regulate, and control the activities that occur within and through it. In this, Lefebvre emphasizes that space is subject to conflict and contestation over the meanings and values that are represented by and attached to it.

Lefebvre's analysis of the production of space centers on the interplay of everyday experience and interactions within historical modes of production. Forms and arrangements of space (spatial practices) as well as representations of space (maps, designs, plans, signs, symbols, regulations, etc.) and the lived experience of space (spaces of representations) exist for Lefebvre as a dynamic interrelationship that collectively produces space. However, for Lefebvre, under capitalism, representations of space dominate the other two elements in that ownership infers control and regulation, that is, the power to delimit and delineate what actions, activities, social practices, and so on, are permitted or allowed in specific spaces.

For Lefebvre, there is a need to understand all three elements in that form, structure, and function alone cannot provide a comprehensive understanding of space. There is a need to ask who produces what forms/kinds of space, why and for what purposes, and whom are they for? But space as a social product is not only the domain of power. Space is also where the lived experience of everyday life takes place, though its inhabitation and colonization is invested with meanings and values that may conflict with the designed intentions of planners, architects, the state, and so on. Space then is the product of social relations and as such is not only a constraining factor, a structuring element, but also the means by and through which we make sense, understand, and invest meaning in our everyday lives.

Space and Place

There is a need to distinguish between the interrelated concepts of *space* and *place*. Often, the two are viewed and used interchangeably, but there is a need for more precision in understanding spatial theories and their application to the analysis of social life. *Space* refers to the abstract conception of physical location and its contents. Spaces become places when individuals and groups assign meanings and social significance to them. Michel de Certeau (1925–1986) argued that space is the effect produced by those processes that seek to define, functionalize, and situate it, while place signifies attachment through uses and practices of meanings and values that confer stasis, value, and importance in everyday life. Similarly, the geographer Yi Fu Tuan (1930–) associates space with mobility, movement, and flow, while place is endowed with values accomplished by pause and immobility through knowledge and use. Thus, there is a distinction between abstract designs and plans, and lived experiences, in which meanings and values change space into place through everyday practices and uses.

David Harvey and the Political Economics of Space

David Harvey (1935–), geographer and influential social theorist, has extended aspects of Lefebvre's thesis on the social construction or production of space as a means to expand Marx's historical

materialism to include a geographical or spatial element. Harvey's initial concentration on the creative destruction and reconstruction of the built environment emphasizes the conflict and contestation over space, which involves not only class conflicts over the allocation and distribution of services and resources but also competition and cooperation between capitalists (through the state) in a profit-oriented market. What Harvey asserts is that space is capable of being shaped and formed by human actions and that such actions do not belong solely to those with power. Command over space involves conflict and contestation over meanings and between exchange and use values. In later work, Harvey extends his analysis beyond the city to focus on the produced space of globalization, particularly that dominated by neoliberal economics, which creates an interrelated network of production, circulation, and consumption of services and commodities, requiring the creative destruction and production of a new global spatial order.

Foucault: Space, Knowledge, and Power

Michel Foucault (1926–1984) did not propose a general theory of space, but his well-known proposition concerning the emergence of a dispersed system of spatial sciences in Europe in the 18th century resulted in what he considered to be the formation of modern disciplinary society. Most of his focus in this respect has been on those technologies of surveillance, in particular the celebrated *Panopiticon* (of Jeremy Bentham), in which architecture and design were applied to the carceral space of prisons to discipline and train prisoners as well as to inculcate in them appropriate moral and physical norms and values that would ensure their rehabilitation before their return to society. The application of architecture and design in which the "eye of power" was used to create "docile bodies" was, according to Foucault, used in other disciplinary institutions, such as schools, hospitals, and barracks. Foucault's writings on space and geography provide a more nuanced understanding of the relationship between power and knowledge as an interlinked concept applied not only in these institutional settings but also to society at large. The organization and control of space for Foucault was a fundamental strategy of bio-politics and "governmentality" in the creation of docile bodies and populations. Foucault singles out the medical profession, in particular public health officers, as the "first specialists of space" who required knowledge of the physical landscape of expanding towns and cities as well as their populations to plan and design social and physical infrastructures to ensure both a more hygienic and a safe urban environment and society.

Spatial theories and analyses have led to an increased interdisciplinary acknowledgment of the importance of spatiality for understanding and research in a variety of spheres and areas. For example, spatial theories have been applied to space and gender (Massey), sexuality (Bell & Valentine), social divisions and fear of crime (Caldeira), postmodern architecture and urbanism (Soja & Gottdiener), landscape (Cosgrove), globalization, and risk.

Andrzej J. L. Zieleniec

See also Durkheim's Philosophy of Social Science; Foucault's Thought; Governmentality and Regime; Human Geography, Social Science of; Postmodernism; Power; Simmel's Philosophy of Society

Further Readings

Bell, D., & Valentine, G. (Eds.). (1995). *Mapping sexual desire: Geographies of sexualities.* London, England: Routledge.

Caldeira, T. P. R. (1996). Fortified enclaves: The new urban segregation. *Public Culture, 8,* 329–354.

Cosgrove, D. (1998). *Social formation and symbolic landscape.* Madison: University of Wisconsin Press.

Crampton, J. W., & Elden, S. (2007). *Space, knowledge and power: Foucault and geography.* Aldershot, England: Ashgate.

de Certeau, M. (1984). *The practice of everyday life.* Berkeley: University of California Press.

Frisby, D. (2002). *Georg Simmel.* London, England: Routledge.

Gottdiener, M. (1985). *The social production of urban space.* Austin: University of Texas Press.

Harvey, D. (2006). Space as a keyword. In *Spaces of global capitalism.* London, England: Verso.

Hubbard, P., Kitchen, R., & Valentine, G. (2002). *Key thinkers on space and place.* London, England: Sage.

Lefebvre, H. (1991). *The production of space.* Oxford, England: Blackwell.

Lefebvre, H. (2003). *The urban revolution.* Minneapolis: University of Minnesota Press.

Massey, D. (2004). *Space, place and gender.* Cambridge, England: Polity Press.

Soja, E. (2000). *Postmetropolis: Critical studies of cities and regions*. Oxford, England: Blackwell.
Tuan, Y.-F. (1977). *Space and place: The perspective of experience*. Minneapolis: University of Minnesota Press.
Zieleniec, A. (2007). *Space and social theory*. London, England: Sage.

Speech Acts

The study of speech acts arose in part as a challenge to a dominant philosophical assumption that the primary purpose of language is to describe the world. In the middle of the 20th century, philosophers such as Ludwig Wittgenstein, J. L. Austin, Peter Strawson, Paul Grice, Gilbert Ryle, and John Searle began to undermine that assumption by arguing that even a grammatically indicative sentence may be used for many other purposes besides description. Describing is now seen as one of a great variety of things that can be done with words, such as defining, promoting, excommunicating, and promising. This new perspective has enabled students of language from many fields (including philosophy, linguistics, social psychology, computer science, literary theory, and law) to appreciate how language use is ineluctably bound up with intentions, interpersonal relations, and social institutions. The last reminds us that speech acts play a crucial role in our social life and, for some, a constitutive role in creating social institutions. In the case of Searle, in particular, something akin to the functioning of speech acts has been put forward as the foundation of the construction of social reality. Speech acts have thus occupied a crucial position in some theories in the philosophy of social sciences.

Speech Acts, Acts of Speech, and Performatives

Many types of action can be performed by the uttering of words. I might startle you with my loud "Boo!" or offend you with my atrocious singing. However, these acts of speech are to be distinguished from speech acts in the semitechnical way in which this expression is used in the philosophy of language. According to that usage, a *speech act* is any act that can be performed by, or in, saying that one is doing so. Growing an inch taller is not a speech act because I cannot do so by, or in, saying so, and likewise for adding a sum to the amount in my checking account. However, I can promise to meet you tomorrow at noon by saying, "I promise to meet you tomorrow at noon," and if appropriately empowered, I can declare the proceedings open by saying, "I declare the proceedings open." Part of the interest of the topic of speech acts derives from our ability to achieve momentous nonverbal effects by verbal means.

Theorists of speech acts typically distinguish among levels of linguistic analysis. On the semantic level, one can string together meaningful words according to grammatical rules to produce a meaningful sentence. "The proceedings are open" is such a case, as is "I promise to meet you tomorrow." Both sentences have fairly determinate meanings, but those meanings are independent of their use to perform a speech act. After all, I might utter either of these in my sleep or while delivering the lines in the play in which I am acting. In neither case do I literally mean what I say (declare any proceedings open or promise to meet you). Thus, uttering a sentence with a semantic content will not ensure that it has been used to perform a speech act. Since the theory of speech acts is a part of the *pragmatics* of language, this is to say that semantic meaning leaves pragmatics underdetermined. A full account of the import of an utterance must attend not just to its semantic meaning but also to its pragmatic significance.

Sometimes, as in the last two examples of promising and declaring open, we make explicit what speech act we are performing. In such cases, we use sentences that one of the pioneers in theorizing about speech acts, J. L. Austin, called *performatives*. On this usage, a *performative sentence* is any sentence that makes explicit the speech act it is characteristically used to perform. Speakers sometimes additionally insert "hereby" to draw attention to the fact that their utterance is accompanied by carrying out the act, as in "I hereby declare the proceedings open."

One can perform a speech act by using a performative sentence. However, we also perform speech acts without making explicit what act we are performing. One can assert that it is raining by saying, "I assert that it is raining," but one can also do so by simply saying, "It is raining." In the latter case, the speaker relies upon the context of her utterance to make it clear that her act is an assertion rather than some other kind. So too in saying, "I'll be there tomorrow," one might count upon context to ensure that

one's utterance is a prediction rather than a promise. However, when something of value is on the line, it can help to make explicit what speech act one is performing with the use of a performative sentence.

Infelicities

Only those appropriately empowered can declare proceedings open. More generally, many speech acts can only be performed by speakers possessing a certain status. Lacking that status, the speaker will perform an act of speech but no speech act—a case Austin termed a *misfire*. For instance, strolling up to the Taj Mahal, I announce, "I hereby declare this the Babe Ruth Palace." Here, I have uttered words with determinate meaning, and I've performed an act of speech but no speech act: I have not endowed that palace with a new name. Austin and his student Searle have detailed the conditions that various acts must meet if they are not to misfire.

Among the conditions that must be met to avoid misfire is the requirement of an appropriate response from the addressee. I might try to bet you $100 that it will rain tomorrow, but I won't succeed unless you accept that bet. Responses such as these are known as *uptake*. Some, but not all, speech acts will misfire unless they are followed by an appropriate uptake.

Austin's doctrine of infelicities in speech acts included not just ways in which they may misfire but also ways in which they may be abused without misfiring. I make a promise I have no intention of keeping. In that case, I have committed myself to carrying out the promised act even though I do not plan to do so. So too, a child might assert that he's done his homework, although he has not done so; here too, a speech act of asserting has been performed, but its insincerity makes it infelicitous. Given the importance for social life of our ability to make agreements, convey information, and modify one another's roles by means of speech acts, we find a common temptation to abuse these institutions for our own gain.

Illocutions, Perlocutions, and Implicature

Austin's term for speech acts was *illocution,* and he distinguished illocutions from what he termed *perlocutions,* which are those effects characteristically brought about by illocutions. Persuading does not fit our definition of a speech act given above, but a characteristic effect of an act of *urging* you to leave the room is persuading you to leave the room. Persuading is thus a perlocution. So too, I cannot impress you by saying that I am doing so, but I may do so with a series of eloquent promises, assertions, appointments, and even denunciations.

Perlocutions are generally effects on addressees. However, speakers often achieve more in a speech act than a perlocution. Often we are able to "mean more than we say" by virtue of the phenomenon known as *implicature* (described and analyzed by Grice). In answer to your question "Where is Susan?" I reply that she is either in the toolshed or in the garage; in so doing, I convey that I do not know which of these two it is. Or you ask me whether I can join you for a movie this evening, and I reply that I have a major deadline tomorrow. I mean, without saying so, that I cannot attend the movie. Grice posited a number of principles (enjoining speakers to be relevant, truthful, adequately informative, etc.) that, as rational interlocutors, we implicitly follow, and he used these principles to explain how we can mean more than we explicitly say. In some cases, this further meaning is a speech act as well. For instance, my remark "You're standing on my foot" is naturally construed not just as a statement of fact but also as a request that you remove your foot from mine. Implicated speech acts such as this are known as *indirect speech acts* (discussed by Searle).

Speech Acts and Norms

Language relates to reality in many more ways than simply describing it, and when it is used to describe reality, it can do so by various illocutionary means. We may appreciate this by noting that different types of speech act are governed by distinctive sets of norms. It is a norm of assertion that one who *asserts* a proposition P should be in a position to respond properly to challenges of the form "How do you know?" By contrast, one might *conjecture* P; and while both assertions and conjectures of P stand to be right or wrong depending on whether P is true, it is not appropriate to call out a conjecture with a challenge to the speaker's knowledge. Rather, an appropriate challenge to a conjecture would be a request that the speaker give *some* basis for what she says; another appropriate challenge would simply be to show that what has been conjectured is untrue. Other speech acts might have the same propositional content P while being governed by yet different norms: Guesses, presumptions, suppositions for the sake of argument, and presuppositions are examples.

Norms governing speech acts evidently overlap with moral norms but are not reducible to them. The injunction that one who asserts that P is to believe that P on pain of being a liar is likely a moral norm. However, one who attempts but fails to name a famous building after himself has done something silly rather than immoral. Norms governing speech acts are perhaps best construed analogously to norms of etiquette. In one society, the "done thing" is to burp loudly after a meal, while in another, this is considered inappropriate. So too, if one is to *assert* or *promise*, certain standards are to be adhered to, but this leaves open the possibility of another culture or group developing its own set of verbal institutions with a unique set of norms.

A twinkle in an eye can speak volumes among sensitive interlocutors. So too, an implication of the theory of speech acts is that the utterance of an innocuous-seeming sentence can be momentous, committing its speaker to a policy or course of action, modifying the social status of the addressee, or forging a bond, for good or ill, between the two. Study of speech acts has and will continue to illuminate the rich communicative content thriving under the grammatical surface. Speech acts are therefore a crucial element of the social world.

Mitchell Green

See also Language, Philosophy of; Language and Society; Searle and the Construction of Social Reality; Semantics and Pragmatics; Sociolinguistics

Further Readings

Alston, W. (2000). *Illocutionary acts and sentence meaning*. Ithaca, NY: Cornell University Press.

Armstrong, D. (1971). Meaning and communication. *Philosophical Review, 80,* 427–447.

Austin, J. L. (1962). *How to do things with words* (2nd ed.; J. O. Urmson & M. Sbisá, Eds.). Cambridge, MA: Harvard University Press.

Bach, K., & Harnish, R. (1979). *Linguistic communication and speech acts*. Cambridge: MIT Press.

Clark, H. (1996). *Using language*. Cambridge, England: Cambridge University Press.

Fish, S. (1976). How to do things with Austin and Searle: Speech act theory and literary criticism. *Modern Language Notes, 91,* 983–1025.

Frege, G. (1984). The thought: A logical inquiry. In *Collected papers on mathematics logic and philosophy* (B. McGuinness, Ed.; M. Black, V. Dudman, P. Geach, H. Kaal, & E.-H. W. Kluge, Trans., pp. 58–77). Oxford, England: Blackwell.

Green, M. (in press). Assertions. In M. Sbisà & K. Turner (Eds.), *Handbook of pragmatics*. Berlin, Germany: Mouton/de Gruyter.

Green, M. (2007). *Self-expression*. Oxford, England: Oxford University Press.

Grice, H. P. (1991). Logic and conversation. In *Studies in the way of words* (pp. 22–40). Cambridge, MA: Harvard University Press.

Hornsby, J., & Langton, R. (1998). Free speech and illocution. *Legal Theory, 4,* 21–37.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic, 8,* 339–359.

Recanati, F. (1987). *Meaning and force: The pragmatics of performative utterances*. Cambridge, England: Cambridge University Press.

Searle, J. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, England: Cambridge University Press.

Searle, J. (1975). A taxonomy of illocutionary acts. In K. Gunderson (Ed.), *Language, mind, and knowledge* (pp. 344–369). Minneapolis: University of Minnesota Press.

Searle, J. (1979). Indirect speech acts. In *Expression and meaning: Studies in the theory of speech acts* (pp. 30–57). Cambridge, England: Cambridge University Press.

Searle, J., & Vanderveken, D. (1985). *Foundations of illocutionary logic*. Cambridge, England: Cambridge University Press.

Strawson, P. (1950). On referring. *Mind, 59,* 320–344.

Strawson, P. (1964). Intention and convention in speech acts. *Philosophical Review, 73,* 439–460.

Tsohatzidis, S. L. (Ed.). (1994). *Foundations of speech act theory: Philosophical and linguistic perspectives*. London, England: Routledge.

Vanderveken, D. (1990). *Meaning and speech acts* (2 vols.). Cambridge, England: Cambridge University Press.

Spontaneous Order

A spontaneous order refers to a pattern, regularity, or structure generated without conscious design or explicit agreement. Sometimes referred to as self-organizing systems, complex adaptive systems, or emergent, endogenous, or polycentric orders, these phenomena may be found in both the natural and the social realms. For example, a particular ecosystem may reflect spontaneous ordering forces within

nature. Such orders may also be found within societies or across societies.

Examples of spontaneous social orders include the emergence of money, social conventions and norms, language, as well as economic patterns and institutions (including, e.g., the division of labor and prices). That a complex ordering might emerge without design, intention, or agreement requires explanation. Social science may explain such orders by using "invisible hand" explanations to explore how a complex pattern might arise from initial conditions in which individuals, bearing certain properties, act so as to bring about the pattern without either intention or design. It is often argued that spontaneous orders coordinate disparate knowledge and generate a complex set of interactions not otherwise attainable by design.

In this entry, we characterize the nature of a social order, discuss how it may be spontaneous, and set forth some of its significant implications.

Order and Spontaneity

To refer to an "order" of things, elements, particles, or persons is to suggest that these manifest relations allow one to make predictions from one part of the order to another. Thus, an order is constituted by relations among parts or elements, a subset of which allows one to form predictions or expectations about other parts or elements. A *social* pattern or order would, therefore, be a complex pattern of actions of human agents whose relations allow a participant to form predictions or reliable expectations about elements of the whole not currently experienced by that person. Within society, the agents may be understood as individuals or as groups (institutions, organizations, firms, etc.). Such a social order is spontaneous insofar as the set of relations emerges, without design or intention, from the actions and interactions of individuals. A spontaneous order is framed by rules, either explicit or implicit. The framework of rules allows agents to respond to and adapt to one another and to changing circumstances; as the agents act on their knowledge (or information) and react to the activities of other agents, new patterns of conduct emerge that were not originally present or part of the intentions of the acting individuals. Such an order may emerge over a relatively short period of time (e.g., an aggregative pattern of market prices), or it may occur over a long period of time (an accumulative order, as exemplified by structures or institutions that evolve over time).

Relevance to Social Science

A number of thinkers have appealed to the idea of spontaneous order. Michael Polanyi was one of the first to employ the terminology, describing the enterprise of science as one such order. In various works, Friedrich A. Hayek has articulated the significance of spontaneous order to an understanding of liberal and commercial societies. For Hayek, a genuinely spontaneous social order emerges out of purpose-independent rules, universally applicable, that provide for plural or private property. Out of these rules, an extended social order will emerge, to include the order of the market as well as overlapping orders, such as those of morals and social norms. The overall spontaneous order (and particularly the market order) will allow for the generation and coordination of more knowledge and information than could have been devised and implemented by a centralized authority or designer. For Hayek and other advocates, a spontaneous order enables individuals to generate and utilize more knowledge, both propositional and practical, than would otherwise be attainable.

Other important social theories bear resemblances to spontaneous order theories. The theory of symbolic interaction, following from the work of George Herbert Mead, suggests that social phenomena emerge as individuals respond to situations, circumstances, and other persons. Recent elaborations of systems theory, in particular the self-reproducing or *autopoietic* theory of Niklas Luhmann, explore how society incorporates a collection of differentiated systems—say legal, economic, or artistic—that function in a decentralized manner.

Relevance to the Philosophy of Social Science

The theory of spontaneous order sets forth salient issues for the philosophy of social science. The very idea of a spontaneous order raises questions of social ontology. Is the order of society to be understood in terms of the behavior of persons? Or is order, in fact, a phenomenon of the mind, constituted by the beliefs and expectations of those who navigate and act within the society itself? The idea

of spontaneous order also broaches significant issues regarding the coordination and utilization of dispersed knowledge. Insofar as such orders allow for a degree and kind of complexity not otherwise attainable, then so should models of regulation and public policy take into account how ever-more precise and determinate rules and regulations may serve to inhibit the growth and utilization of knowledge and information.

Eugene Heath

See also Austrian Economics; Complexity and the Social Sciences; Cultural Evolution; Emergence; Emergence and Social Collectivism; Hayek and the "Use of Knowledge in Society"; Invisible Hand Explanations; Symbolic Interactionism; Systems Theory

Further Readings

Buchanan, J., & Vanberg, V. J. (1991). The market as a creative process. *Economics and Philosophy, 7*(2), 167–186.

Cantor, P., & Cox, S. (Eds.). (2009). *Literature and the economics of liberty: Spontaneous order in culture.* Auburn, AL: Ludwig von Mises Institute.

diZerega, G. (2008). New directions in emergent order research. *Studies in Emergent Order, 1,* 1–23.

Hamowy, R. (1987). *The Scottish Enlightenment and the theory of spontaneous order.* Carbondale: Southern Illinois University Press.

Hayek, F. A. (1964). Kinds of order in society. *New Individualist Review, 3,* 3–12.

Hayek, F. A. (1973). *Law, legislation, and liberty: Vol. 1. Rules and order.* Chicago, IL: University of Chicago Press.

Kuntz, P. (Ed.). (1968). *The concept of order.* Seattle: University of Washington Press.

Luhmann, N. (2008). *Law as a social system.* Oxford, England: Oxford University Press.

Mead, G. H. (1934). *Mind, self, and society.* Chicago, IL: University of Chicago Press.

Menger, C. (1985). *Investigations into the methods of the social sciences with special reference to economics* (L. Schneider, Ed.). New York, NY: New York University Press. (Original work published 1883)

Polanyi, M. (1998). *The logic of liberty.* Indianapolis, IN: Liberty Fund.

Schelling, T. (1978). *Micromotives and macrobehavior.* New York, NY: W. W. Norton.

Sugden, R. (1989). Spontaneous order. *Journal of Economic Perspectives, 3,* 85–97.

Tuebner, G. (Ed.). (1988). *Autopoietic law: A new approach to law and society.* Berlin, Germany: Walter de Gruyter.

Straussian Critique of Social Science

The Straussian critique of modern social science originates in the writings of Leo Strauss (1899–1973), who sought to restore an appreciation of classical political philosophy as a superior mode of addressing fundamental human questions. This entry reviews Strauss's thought on the primacy of political philosophy, his sustained critique of "new" or "behavioral" political science, and shows its continuing importance for recent developments in political theory.

Strauss criticized the social science of our time as *insufficiently scientific* or precise, as the result of the misguided attempt to apply the methods of modern natural science to phenomena that do not lend themselves to understanding through that approach. Strauss regarded the recovery of the classical perspective as particularly urgent in our era, owing to the crisis of modern Western civilization, which was both articulated and deepened in the thought of Friedrich Nietzsche and Martin Heidegger. Strauss's aim was not to secure agreement on all of the specific political judgments made by the greatest classical writers, such as Plato, Aristotle, and Xenophon—judgments that would certainly need to be revised in light of subsequent political, religious, and scientific developments. Nor did he deny the utility of quantitative scientific methods in a field like economics (the one area of social science that has clearly profited from that approach). But he argued for the *primacy of political science or political philosophy* over the other human sciences (as Aristotle suggested in calling politics the "architectonic" study, or *technê*, situated at the summit of all others and directing or "governing" them all) and for the necessity that the student of society *begin* his inquiries, as the classical philosophers had done, with the perspective of the concerned citizen, who seeks clarification of the nature of his own and his country's good and makes claims on his government based on opinions about justice. It is the fundamental task of the political

philosopher, Strauss maintained, to clarify the issues raised in such civic debates, with a view to both promoting practical political improvement and achieving greater understanding of the human condition itself.

The chief obstacle to social science's furthering such goals in our time, Strauss contended, was the supposed distinction between "facts" and "values," developed most profoundly by the great German sociologist Max Weber and adopted by the 20th-century theorists of "logical positivism." According to that distinction, it is the duty of the social scientist to limit his judgments to matters of fact, eschewing evaluative claims, since the latter are incapable of being verified by reason. In the second chapter of his most influential book, *Natural Right and History,* Strauss challenged Weber's argument, contending that it is impossible to provide meaningful, empirical descriptions of social or political phenomena without evaluating them; that the supposed incapacity of reason to arrive at objective evaluations is itself incapable of being demonstrated; and that the fact/value distinction, contrary to Weber's intention, inevitably leads to outright moral relativism, or nihilism, according to which the basest of human deeds must be accorded "objectively" equal status with the noblest.

Although Strauss addressed the problems of social-science relativism in several of his other writings, his most comprehensive critique, outside of the chapter in *Natural Right and History,* was the "Epilogue" he composed to *Essays on the Scientific Study of Politics,* edited by Herbert Storing, in which four of his former students offered critical assessments of leading examples of the "behavioral" approach to different areas of contemporary political inquiry: voting studies, public administration theory, group politics, and "scientific propaganda." Summarizing and elaborating the findings of those studies, Strauss identified several defects of the "new" political science, including the following.

First, despite their profession of relying only on information about politics derived from "scientifically" verifiable research, "behavioral" political scientists are "constantly compelled to borrow from common sense knowledge" both to identify and understand objects of knowledge and in order to make their findings intelligible—thus attesting to the fact "that there is genuine prescientific knowledge of political things" possessed by the unscientific citizen that "is the basis of all scientific knowledge of them" (Strauss, 1968, p. 214).

Second, in its attempt to uncover "universals" that are assumed to be common to all political regimes, the "new" political science obscures the true universals in light of which political phenomena need to be understood: the differences *among* political regimes, which in turn shape the character of subpolitical associations. Thus, it misleadingly reduces distinctions of kind (e.g., between liberal democracy and communism) to mere differences of degree.

Third, because the chief techniques of contemporary, "empirical" social science research can be applied only to human beings now living, in countries where governments tolerate such research, its supposedly universal findings tend to be parochial in both time and place.

Fourth, in its attempt to replace "the political understanding of political things" with a "scientific" language, such as the substitution of "power relations" for "politics" and of Sigmund Freud's "superego" for conscience, behavioral political science obscures crucial distinctions (e.g., between a bad conscience and "guilt feelings") as well as the more general complexity of political life that is captured in ordinary speech.

Fifth, and most important, having ostensibly expunged "values" from their descriptions of political things, the new political scientists are compelled to smuggle evaluations in through the back door (e.g., through terminology like "the open society"); but unlike the practice of the classical political philosophers or intelligent modern historians, these evaluations are never subjected to rational assessment. Hence, they remain mere prejudices.

Although fads in the social sciences have partly changed since Strauss wrote, his criticisms apply *mutatis mutandis* to more recent movements like "rational choice theory" and the "new normativism," typified by the writings of John Rawls and Robert Nozick, which constructs utopias ungrounded in political fact or in critical assessment of the author's value preferences. While Strauss recognized that a considerable amount of valuable work is done by genuinely empirical political scientists—on topics such as the operation of American political institutions or problems of contemporary international relations—he tried to rescue such practical political analysis from subsumption under abstract theories that obscure rather than

clarify the nature of political phenomena. Without denying the achievement of the great modern political philosophers whose thought generated America's constitutional-liberal regime, he maintained that Aristotelian political philosophy—which bears certain resemblances to the political science of Alexis de Tocqueville—offers the best theoretical foundation for political inquiry.

David Lewis Schaefer

See also Behavioralism in Political Science; Common Sense (in the Social Sciences); Logical Positivism/ Logical Empiricism; Naturalism in Social Science; Objectivity; Philosophy of Politics, History of; Value Neutrality in Science; Weber and Social Science: Methodological Precepts

Further Readings

Behnegar, N. (2003). *Leo Strauss, Max Weber, and the scientific study of politics.* Chicago, IL: University of Chicago Press.

Pangle, T. L. (2006). *Leo Strauss: An introduction to his thought and intellectual legacy.* Baltimore, MD: Johns Hopkins University Press.

Schaefer, D. L. (1991). Leo Strauss and American democracy. *Review of Politics, 53*(1), 187–199. (Reprinted in *Leo Strauss: Political philosopher and Jewish thinker,* K. L. Deutsch & W. Nicgorski, Eds., 1994, Lanham, MD: Rowman & Littlefield)

Storing, H. J. (Ed.). (1962). *Essays on the scientific study of politics.* New York, NY: Holt, Rinehart & Winston.

Strauss, L. (1953). *Natural right and history.* Chicago, IL: University of Chicago Press.

Strauss, L. (1978). *Liberalism: Ancient and modern* (Foreword by A. Bloom). New York, NY: Basic Books. (Original work published 1968, Chicago, IL: University of Chicago Press)

Strong Program in the Sociology of Scientific Knowledge

This entry explains the basic theses of the influential "Strong Program" in the sociology of scientific knowledge, also known as the "Edinburgh school," whose main protagonists were Barry Barnes and David Bloor.

Supporters of the Strong Program argue that both true and false, and rational and irrational beliefs are open to sociological study and that the same general explanatory processes and causes are at work in the analysis of both cases. The principle that sociological curiosity should apply equally to all knowledge claims is called the *symmetry postulate*.

The Strong Program is best approached via the position that its supporters reject—a position that may be called the "weak" program. The weak program is based on the idea that sociological causes may explain (a) the conditions that promote or inhibit science and (b) any deviations from the operation of the rational principles that should govern the growth of knowledge. Sociologists can be given the task of explaining why science developed in Europe and not in China or the job of explaining why science is distorted by political interference in totalitarian states, but that is all. According to the weak program, the normal, rational activity of science calls for a nonsociological explanation in terms of the reasons advanced by scientists themselves.

The weak program rests on an allegedly fundamental distinction between the *rational* and the *social*. Society can, at most, propagate rationality but cannot constitute it. Society is the cause of facilitation or deviation but not the immanent working of rationality itself. The weak program has clear implications for history and sociology: It is necessary to disentangle the social from the rational and exhibit each playing its proper, but different, role. The attractions of this view may derive from its roots in daily life. What is normal and routine becomes taken for granted, and only deviations seem to call for explanation. Commuters want to know why the train is late, not why it is on time. In reality, there is a cause for the train being on time just as much as for its being late. This may arouse no interest in the tired traveler, though it should be highly interesting for experts who want to keep the system running. The weak program, as formulated, for example, by Imre Lakatos, involves turning this commonsense asymmetry and lack of curiosity about routine into a philosophical principle.

The weak program may be superficially attractive, but there is a price to be paid. The most cogent criticism of the weak program derives from Ludwig Wittgenstein's analysis of rule following. He took the example of following an arithmetical rule: for example, "add 2" in order to generate the sequence

2, 4, 6, 8, and so on. Critics of the Strong Program argue as follows: Surely the behavior is governed by the rule, and as long as the rule is followed, the rule itself explains the behavior. What is there for the sociologist or psychologist to explain except (a) how the follower came to be in a position to understand the rule (e.g., schooling) and (b) deviations from the rule (e.g., lapses of attention or inadequate training)? It should be clear that on this view the operation of the rule "itself" remains mysterious.

Wittgenstein sought to provide a more revealing "strong" account of rules by examining the resources available when training a rule follower or when correcting errors in performance. He found, of course, that reasons come to an end. Any correction offered by a teacher will be a rule for following a rule or an interpretation of the rule. To avoid a regress, a tenable explanation must ultimately depend on there being ways to follow rules that operate below the level of reason giving. The process ultimately depends on ways of aligning behavior at the causal level. Everything depends, as Wittgenstein put it, on something animal. We are social animals susceptible to one another through the influences generated by our interactions. Thus, *the rational depends on the social* and cannot be disentangled from it. This amounts to accepting that rationality is indeed a natural phenomenon (nothing supernatural about it) and that rules are conventions and social institutions. Wittgenstein's work thus destroys the most cogent supporting example of the weak program. It shows that there is an *irreducible* sociological dimension to correct rule following, not just to incorrect rule following, just as there is a sociological dimension to trains that run on time.

Two further points need to be stressed. First, the Strong Program is not just concerned with examples of the kind given above. It also applies to empirical knowledge. Contrary to the claims of critics, the program does not imply that knowledge arises from scientists responding to society *rather than* to the material environment. Scientific knowledge of the material world is indeed about an independent reality, but the knowledge is always shared and collectively constructed. As Thomas Kuhn demonstrated in his *Structure of Scientific Revolutions,* these sociological characteristics apply to good science just as much as to bad science. Notice, though, that it is *knowledge* that is socially constructed, *not* the ultimate constituents of the material world.

Second, the Strong Program is *relativist*. Relativism simply means the rejection of any claim to absolute knowledge and derives from an awareness of the natural origins of all processes of cognition. Relativism must not be confused with idealism ("the material world does not exist"), irrationalism ("anything goes"), or the idea that humans are imprisoned in inherited conceptual schemes ("the myth of the framework"). As Wittgenstein's arguments show, conceptual schemes (like rules) are the product of social life, not its determinants, just as the shared use of a word determines its meaning, rather than meaning determining use.

David Bloor

See also Feyerabend, Critique of Rationality in Science; Kuhn and Social Science; Lakatos, Methodology of Scientific Research Programs; Language-Games and Forms of Life; Relativism in Scientific Theories; Relativisms and Their Ontologies; Rule Following; Science and Ideology; Social Constructivism; Social Studies of Science and Technology; Sociology of Knowledge and Science

Further Readings

Bloor, D. (1976). *Knowledge and social imagery.* London, England: Routledge.

Bloor, D. (2011). Relativism and the sociology of scientific knowledge. In S. Hales (Ed.), *A companion to relativism* (pp. 433–455). Oxford, England: Blackwell.

Kuhn, T. (1962). *The structure of scientific revolutions.* Chicago, IL: University of Chicago Press.

Lakatos, I. (1971). The history of science and its rational reconstructions. In R. Buck & R. Cohen (Eds.), *Boston studies in the philosophy of science* (Vol. 8., pp. 91–136). Dordrecht, Netherlands: Reidel.

Wittgenstein, L. (1967). *Philosophical investigations.* Oxford, England: Blackwell.

Structural Functionalism, in Social Theory

Structural functionalism became one of the most influential theories in the history of sociology and also figured quite significantly in that of anthropology. Structural functionalism is not to be confused

with *structuralism,* for which there is a separate entry in this encyclopedia.

Origins

Structural functionalism originated in the work of some of the founders of sociology—Auguste Comte (1798–1857), Herbert Spencer (1820–1903), and Émile Durkheim (1858–1917). It reflected their interest in seeing society as in many ways being analogous to a biological *organism,* with its separate parts each fulfilling necessary *functions.* So just as the stomach performs a digestive function and the heart the function of pumping blood about the body, so in the social body the family would fulfill the functions of procreation and socialization, schools an educational function, and so on.

Of the three, Durkheim's work was by far the most influential. In addition to his sociological influence, his writings profoundly affected the functionalist leanings of *social anthropology* through the work of Alfred Radcliffe-Brown (1881–1955) and Bronislaw Malinowski (1884–1942). Durkheim's own position may be more complicated with regard to his espousal of functionalism, but it may be safe to say that his book *Division of Labor* is the inspiration for functionalism and functionalist interpretations of him. In addition, Radcliffe-Brown has some claim on the title of "founder of structural functionalism" (though he was not happy with the label).

These latter two social anthropologists, through their disagreements over what is perhaps the central concept of structural functionalism—that of *function*—provided deep insightful thought into the philosophical foundations of their discipline and perhaps, ironically, the grounds for later critique.

Radcliffe-Brown, in keeping with his desire for scientificity, wished for conceptual precision with regard to the notion of function. He thought that with regard to the two ways science had used the term—as the physiological contribution a particular organ made to the continued healthy working of an organic system or with the quite different meaning and context of a mathematical function—the former, but not the latter, should be developed analogously. Thus, the notion of function for Radcliffe-Brown should refer to the contributions made to the persistence of the entire system of social relations. Malinowski criticized this both as unnecessary abstract formalism and as problematic in terms of the role of analogies in social science derived from the natural sciences. In this latter point, he presaged quite important thinking in the philosophy of social science, though, at the same time, espousing a slightly naive dependence upon the "concrete" in fieldwork.

One could make the argument, however, that Malinowski won this debate because whatever the theoretical merits or not of Radcliffe-Brown's position, the actual history of usage of the term *function* shows considerable slippage and imprecision, which we shall briefly consider later in this entry.

Neither the early sociologists nor the anthropologists fully developed the systematic body of theory known as structural functionalism. Rather, they should be seen as progenitors, as the theory awaited a later thinker for its full development: Talcott Parsons (1902–1979). Ironically, Parsons was not particularly happy with this as a label for his theory, but be that as it may, it is the label that is indelibly associated with his name.

Talcott Parsons

Sociology is often described as a discipline lacking a paradigm. That is, it lacks a common philosophical foundation, methodology, and dominant mode of theorizing. It even lacks an agreed-upon definition of its subject matter and what its most important questions are. This is truer than ever today; and though this lack of foundational consensus has been a historical constant, it did at one time come close to a paradigm. That was during the 1940s and 1950s and reaching into the early 1960s, when structural functionalism was the dominant sociological perspective, less so in Europe but particularly so in America.

A good example of sociology lacking consensus can be seen in the work of two of its three most important classic thinkers: Max Weber and Émile Durkheim (the other being Karl Marx). Durkheim believed that society had a sui generis reality, whereby it existed in its own right and was not reducible to the sum of its constituent parts. Weber not only disagreed with this perspective but considered it an error—the error of reification—made not only by social theorists but by ordinary people as well. The error of reification, according to Weber, is to conceive of collective entities, say, a team, as really existing as something in their own right. Rather, a

team should be thought of as a convenient linguistic fiction; the reality of the team is the individual people: the coach and players who compose it. Weber's theoretical emphasis was upon social action, while Durkheim focused more upon *social structure*.

It is the latter emphasis that structural functionalism tends toward, but it was Parsons's conscious intent to combine the best insights of *both* Weber and Durkheim. We can see this intention in the title of one of his most famous works, *The Structure of Social Action*. We can also see that structural functionalism leans toward the structure side of the equation from the title of another of his most famous works, *The Social System*.

The first step in the scientific investigation of any social phenomenon, Parsons argued, is the identification of the *unit acts* of which it is composed. The unit act is the most fundamental concept of all social analysis and is analogous to an elementary particle in physics. The second step is to describe their features in terms of the *action frame of reference*. This regards actions as being organized over time and involves the purposive use by people of means to attain goals.

From this perspective, we can see our social system as being an action system and, thus, a system of interrelated parts. These parts *function* in relation to the needs of the whole that is society. These parts not only perform some very specific functions, but, as Parsons argues, every element also contributes to the stability of the system as a whole. In this regard, we can see Parsons's thought as being consistent with that of Radcliffe-Brown.

In order to maintain order and reproduce themselves, societies have *functional prerequisites*. They all have fundamental needs that must be fulfilled. Any action system (that lasts) will have these, and only these, four main functions successfully fulfilled: (1) a set of adaptive processes, (2) a means for attaining goals, (3) a set of institutions performing an integrative function, and, last, (4) the somewhat oddly named *latent pattern maintenance* function. These fundamental functional requirements form the famous acronym AGIL, through which students have been memorizing Parsons's ideas for generations: *A* is for adaptive function, *G* for goal attainment, *I* for integrative function, and *L* for latent pattern maintenance. The following discussion examines each of these in turn. But, first, we must note here the previously mentioned conceptual slippage of the term *function*. Are these "functions" need-fulfilling ends in themselves for Parsons? It would seem that something more akin to fundamental subsystems would be the intended meaning.

First, the social system must adapt itself to its external environment, its physical environment. This *adaptive function* is performed (primarily) by the various economic institutions. The economy *adapts* the environment to our social needs through the production and distribution of goods and services.

Second, society needs institutions and processes whereby people can be mobilized to attempt to attain collective goals. This *goal attainment* function is principally fulfilled through political institutions. By this is meant political institutions in the broadest of all possible senses. Trade unions, for example, are perfect examples of such institutions.

Third, the *integrative function* is fulfilled by the institutions and processes whereby social cohesion and solidarity are achieved. These are the institutions that allow for the coordination of different activities and functional units. This coordination, Parsons believed, very importantly included principles of social stratification. The people with the right set of skills and abilities needed to be recruited to the appropriate jobs.

Finally, there are the institutions and processes that function to ensure the continuity of fundamental normative values. The processes whereby this is achieved are usually latent and underneath the surface, rather than explicit (hence the word *latent* in the term *latent pattern maintenance*). These processes are found in the institutions that socialize us (e.g., the family and the education system). Such processes are among *the* most crucial.

The social system is a *dynamic* system, and though always tending toward stability, it nonetheless evolves. An adaptive-upgrading process thus accompanies society's evolution. The theory assumes that the new subsystems that develop will be more adaptive than the earlier ones. Society becomes more and more differentiated, particularly in the economic realm. This differentiation leads to new problems of integration and creates new needs—and the corresponding development of appropriate skills and abilities. New subsystems thus emerge.

An important way of conceiving of society as a social system is to see it not as a collection of

individuals but rather as an interlinked system of positions. Each position possesses one or more functions and has *roles* corresponding with them that the individuals who occupy the positions must fulfill. Any given individual will have a number of different roles according to the different positions he or she occupies in the system. Thus, a woman may be both mother and daughter, with correspondingly different roles for each position. A person may be a student *and* an employee. And so on. There are myriad positions and roles possible, limited only by the complexity of the particular society.

Some of the roles are permanent, even though they may change over time. For example, a daughter's responsibilities change as she gets older, until perhaps later on in life she might assume a caregiver's role with respect to her own mother. Many other roles are by their very nature temporary. For example, Parsons describes what he called the "sick role." When we are sick, this discharges us from the usual responsibilities of our job and domestic chores. However, with this role come responsibilities, the responsibility to heed the doctor's advice, for example, as part of the overarching responsibility to get well. It is clear with this example that the concept of roles embodies considerable expectations of behavior from individuals.

Individuals are not generally coerced into conformity with these expectations. Rather, as part of a lifelong process, they are socialized into the roles they occupy. From a structural-functionalist perspective, the *self* is an *evolving collection of positions and roles and responsibilities,* a changing set of patterns of behavior that form a personality. In this way, Parsons demonstrates the important influence of Sigmund Freud upon his work.

Criticism

There have been many criticisms of Parsons. George Caspar Homans, Parsons's colleague at Harvard, felt that Parsons had failed to explain social structure because the overly elaborate and abstract structure did not properly take into account the *actions of individuals*. In their ongoing debate, they thus reproduced to some extent the ontological and epistemological differences of Weber and Durkheim—precisely what Parsons hoped his own work would transcend.

This debate also occurred within the anthropological tradition of functionalism and its critics. Is there a different order of reality beyond the actions of individuals? Given that social structure is not immediately visible, is the empirical study of it possible even if it does exist? The ontological questions of existence translate into epistemological questions as to the possibility of knowledge, which in turn translate into some practical difficulties of methodology. While you reach a dead end with Parsons from this perspective because he always stayed at the level of grand generalizations, Radcliffe-Brown and most particularly Malinowski *did* translate a functionalist perspective into an empirical investigative methodology—only to encounter the same charges of reductionism voiced by another major figure in anthropology, Edward Evans-Pritchard.

There seems to be an implicit *circularity* in the basic premise of functionalism. An institution exists in order to fulfill certain functions—therefore, if it exists, it must be functional. It should be noted here that it was precisely this point that Radcliffe-Brown objected to in some notions of functionalism.

In addition to this circularity, there is also an inherent *ahistoricism*. Parsons seems to have taken the existing American status quo and universalized it as applying to societies everywhere and always. Parsons's theory precludes questions (standardly voiced against structural functionalism) as to whether any other, *different sort of institution* could fulfill the same function as the one carried out by any currently existing institution. The (nuclear) family, for instance, fulfills the functions of procreation and socialization. However, it is apparent that other, very differently organized institutions could as easily fulfill this function.

One of the most sympathetic of Parsons's critics was Robert K. Merton (1910–2003), who also pointed to philosophical contradictions with functionalism. His principal contributions are twofold. First, Merton introduced the distinction between *manifest* and *latent* functions. The former are related to the conscious intentions of actors, and the latter deal with the unintended consequences of action. The actor may not be aware of all the functions of his or her position or what the actual consequences of an action may be in this regard.

Second, Merton argued that while elements of the system sometimes function in relation to preserving

the overall unity of the system (a primary argument for Parsons), this is not always the case. Elements can be functional for some yet highly dysfunctional for others. Unemployment, for example, may economically function to keep wages down, because the greater insecurity about their jobs felt by the workers makes them reluctant to demand improvements. While obviously this is not functional for the unemployed themselves, it may also be doubted that it is functional for society as a whole. Is it really a good thing for the economy that wage demands are kept low? While it could benefit some employers, it is not beneficial to those whose wages are driven down.

Through this sort of example, power and conflict are admitted to the structural-functionalist schema. The emphasis upon social order and social stability inevitably gave way to perspectives of conflict and change.

While there has been *relatively* little sociological interest in structural functionalism and Parsons since the 1970s, there has been a more recent revival, primarily in the Neo-functionalism of Jeffery Alexander (1947–).

Garry Potter

See also Durkheim's Philosophy of Social Science; Feedback Mechanisms and Self-Regulatory Processes in the Social Sciences; Holism, in the Social Sciences; Social Anthropology; Systems Theory; Weber and Social Science: Methodological Precepts

Further Readings

Alexander, J. C. (1988). Parsons' "structure" in American sociology. *Sociological Theory, 6,* 96–102.

Durkheim, É. (1997). *The division of labor in society* (L. A. Cosner, Trans.). New York, NY: Free Press.

Homans, G. C. (1964). Bringing men back. *American Sociological Review, 29,* 809–818.

Malinowski, B. (2001). *A scientific theory of culture and other essays*. New York, NY: Routledge.

Merton, R. K. (1949). *Social theory and social structure.* New York, NY: Free Press.

Parsons, T. (1951). *The social system.* Glencoe, IL: Free Press.

Parsons, T. (1967). *The structure of social action: Vol. 1. Marshall, Pareto, Durkheim*. New York, NY: Free Press.

Radcliffe-Brown, A. R. (1952). *Structure and function in primitive society: Essays and addresses*. London, England: Cohen & West.

Structuralism and Poststructuralism

Structuralism and, subsequently, poststructuralism were key components in what became known as the "linguistic turn" in 20th-century social theory. Both are less accurately described as theories; they are, rather, broadscale theoretical orientations, each of which served as a foundation for a rich and diverse number of thinkers and researchers.

This entry provides an overview of structuralism and poststructuralism and how they relate to both social science and philosophy.

Structuralism

Structuralism originated in the work of Ferdinand de Saussure (1857–1913). Seldom in the history of thought has there been a single book so influential as *Course in General Linguistics* (published posthumously by his students from their course notes). With this work, the *synchronic* study of language and meaning generation truly began, and Saussure is thus credited as the "father of modern linguistics." Previously, language was mainly studied diachronically; that is, the evolution of words and grammatical construction, and so on, was traced over time.

It is difficult to overstate the revolutionary nature of Saussure's work. His synchronic study of language focused upon the manner in which meaning is generated *structurally* through an arbitrary "sign" system of differences (as described below). The realization that meaning is structurally generated through *sign systems* took his linguistic analysis beyond the study of the spoken and written word to embrace potentially *all* sign systems (of which our spoken and written language is merely a subset) involved in human communication. Thus, an entire new discipline was born—*semiology*—and once again Saussure is credited as the "father" of it.

Through semiology, the entirety of human culture—and each and every aspect of it, from architecture to literary criticism to the unconscious mind—can be analyzed in terms of its being a structured sign system. A plethora of famous thinkers (though many did not accept the structuralist label) applied this methodology to a diverse range of human phenomena, for example, Roman Jakobson

(linguistics), Claude Lévi-Strauss (social anthropology), and Jacques Lacan (psychoanalysis). Louis Althusser (Marxism) and Roland Barthes (literary criticism and cultural analysis) are perhaps the most famous.

The technical fundamentals of semiology are fairly simple. Language (or any meaning-generating sign system) can be broken down into two categories: *langue* and *parole* (or in English, "language" and "speech"). Parole is what is happening for me now as I type this and for you as you read it. When in speaking you utter a command or ask someone a question, that is parole, too. When they hear and understand you, that is also parole. One way of putting it is to say that parole (or speech) is any *particular* usage of language. Saussure's great insight was to realize that for *any* particular instance of language usage to potentially work in terms of intended and received meanings it must depend on the underlying system of linguistic rules or structure of language, that is, on *langue*.

A sign is composed of two elements, a *signifier* and a *signified,* which fit together as the two sides of the same coin, of the sign itself. In written or spoken language, the signifier is the written mark or sound, respectively, and the signified is the mental image associated with the mark or sound.

A more naive view would be to see the essential meaning-generating relationship as between a word and a thing. So there is something out in the world, a chair, for instance, and we label it with a sound or set of marks: *c h a i r.* But this is a highly inadequate theory of meaning. For example, understanding the relationship between particular entities and categories of entities would be doomed to confusion with such a naive, pointing-and-labeling theory of meaning. The signified is *not* any real-world referent but rather the *image in the mind* (of both speaker and listener, writer and reader) provoked by the signifier. Exactly where real-word referents stand in this understanding of meaning and language is discussed later in this entry.

Crucial to this theory of meaning is the thesis that the relationship between signifier and signified is *arbitrary.* What is meant by this can be most clearly demonstrated by a consideration of nonarbitrary examples. Onomatopoeia is one such case. The word *moo,* meaning in English the sound a cow makes, has some similarity to the actual sound a cow makes; though as Saussure wittily pointed out, apparently the sounds French and English animals make must be different as they have different words for these sounds. Another case would be pictogram writing. In such forms of writing, the symbols resemble, usually in highly stylized ways, what they mean. An example would be a pictogram resembling a house and meaning the same.

Saussure stresses that such cases are the *exception*s to the crucial rule of meaning and sign systems—namely, that the relationship between signifier and signified is *arbitrary.* The very importance of this assertion warrants its repetition. What is meant by this is often illustrated in the famous example of the "cat" and the "mat." What makes the former signify a four-legged feline creature and the latter something you wipe your feet on is nothing more than the fact that a *c* is not an *m*. The written alphabet is nothing more than a system of arbitrary differences. So too is the sound system of "c" and "m" sounds.

Meaning is generated through a series of "syntagmatic" and "paradigmatic" choices. Structural analysis in semiotics takes place along two axes of meaning: On the horizontal plane, a *syntagm* is a combination or concatenation (chain) of several successive linguistic units of the same level, thus forming a meaningful whole, for example, a sentence, while on the vertical plane a *paradigm* is a class of linguistic elements grouped together according to a certain principle. Signs make sense—have meaning—only because they are situated within the structure of these two axes or planes, that is, they are nodes in a system or structure, namely, the system of language.

So, as we are speaking the sentence <the cat was asleep on the sofa>, we are choosing the sound signifiers of the "c," "a," and "t" sounds (combined to form the "cat" sound) and also making a series of other choices. We choose *c* instead of *m* or *b*, *cat* instead of *lion* or *tiger* or *cheetah* or *moose*. We choose the sound order "the cat" rather than "cat the." And so on. The *conventions* of syntax and grammar in language, along with the *arbitrary conventions* of the signifier and signified relationship, are what enables meaning to be produced. The syntactical and grammatical choices are, of course, going to be different in different languages (or any sign system). Foreign learners of English, for example, make different common errors depending

on their native language. They do so because they incorrectly apply in English the grammatical conventions of the other language.

What holds the system together and allows it to generate meaning is thus nothing more than *social convention*. However, that "nothing more" is actually a very great deal theoretically. In this regard, we can plot a trajectory leading back from Saussure to Émile Durkheim's sui generis *social facts*. Language, Durkheim asserted, was one of the most crucial *social facts* and illustrated an independently existing *different order* of reality—a level of reality not reducible to a physical materiality.

Poststructuralism

It is such ontological questions as the preceding that one sees as constituting one of the crucial launching points for the birth of poststructuralism. Another concerns the referent question. That is, if a sign consists of a signifier and signified and the latter is *not* an actually existing object in the world but rather an image in the minds of the message instigator and recipient (that this is the same image is guaranteed by the social conventions of the sign system), then the question arises as to the ontological (existential) status of the referent. When we are talking about a cat, for example, the sound signifier of "cat" provokes in us the same signified image; and if we are talking about a particular cat, "this cat" in the room, it is no less true that what is provoked by the signifier is a signified, an image—not the independently existing (i.e., independent of our discourse) particular animal in the room. It only remains to point out that to say "independently existing" is an ontological position consistent with the philosophical realism that appears to be implicit in structuralism, but *not* with poststructuralism. Saussure himself does not provide a definitive answer to this question; one could classify him as an agnostic in this regard.

While Saussure can rightly be considered the founding father of structuralism (and thus indirectly of poststructuralism), it was left to others to develop it further. While modern linguistics derives from Saussure, it has long since moved on; he would only get a passing mention in a contemporary linguistics course. He has actually probably been more influential on other disciplines.

In social anthropology, for example, the work of Claude Lévi-Straus had considerable impact for a time and intellectually popularized the term *structuralism*. He provided a structuralist analysis of both kinship relations and mythology. His analysis of both these disparate areas of investigation can be seen to directly connect with Saussure. For example, the empirical variety of tribal myths found in the world can be analyzed according to a system of elements combined and recombined to produce different meanings. Similarly, we can see the emphasis on the relational formation of meaning in his perspective on kinship systems. An uncle, a nephew, a brother, a sister and so on, are individual positions in a system and find their significance and prescribed roles not in and of themselves but rather in their *relations* to the other positions in the system.

The most direct intellectual descendant of Saussure, however, is the discipline of semiology as applied to cultural analysis and literary criticism. Later, this was developed in a variety of ways by poststructuralists, and we can find both semiology's structuralist origin and its basis for this later development in the work of Roland Barthes. Barthes made the observation that we frequently find in bourgeois culture (Barthes was a Marxist as well as a structuralist) a variety of *levels* of signification. So on one level perhaps we have a simple denotation. The words *red* and *rose* are the signifiers for our mental image of a particular colored flower. On another level, the signified red rose is itself a signifier in another signifying system: a red rose signifies love.

The example given in the preceding paragraph may seem perhaps rather banal and obvious, and through its very banality thus inherently *innocent*. This is very much the point of the sort of analysis Barthes engaged in. Our culture is precisely *not* innocent. The levels of signification serve to create a false naturalness with regard to the many, extremely politically nuanced meanings it is constantly producing.

Barthes demonstrated the above in a variety of diverse contexts: in literary works, in the language of fashion, and in the spectacle of professional wrestling (one wonders if the founders of today's WWE studied Barthes's work!). However, the structuralist analysis of culture opens up some questions concerning its own assumptions that almost beg for their transcendence. If language, if culture, is a structured system of oppositions, then what guarantees the necessary stability of meaning for successful communication to occur?

We may recall here that the essential arbitrariness of the signifier–signified bond means that at the root of all meaning is merely social convention. But

societies are fluid and changing. It follows that meaning is never absolutely fixed; there is no transcendental signifier to ground it; meaning is also fluid and ever-changing. Ergo, poststructuralism was born. Important contemporary thinkers working within the poststructuralist paradigm are too numerous to mention, but Jacques Derrida and Michel Foucault are seminal original influences.

Derrida embraces the contradictions implicit in structuralist semiology and both celebrates them and puts them to work. He further develops the notions of meaning fluidity and textual convention as the only grounding for meaning. The idea is that there are levels of signification—that every signified is itself a signifier—reaching to infinity, an "endless chain of signification," where there can never be true denotative meaning. What this last thought means is that in order to understand a word requires context—the context of the sentence. But the sentence too requires contextualization—the contextualization of the paragraph, the page, the chapter, the book, the history of thought. So meaning is never fixed but rather interactively produced and reproduced between reader and text.

Commencing with such premises led Derrida to produce not a philosophy or even, as some would have it, an "anti-philosophy" but rather a method of textual practice: *deconstruction*. It is a very particular kind of method and one with profound philosophical implications, but at root it *is* a method. Deconstruction begins with the idea that among the paired oppositions through which much meaning is generated, that between *core* and *peripheral* is of particular importance. It is where the innocence of language is lost and power becomes integral to meaning itself. In Derrida's deconstruction of many of the seminal texts of Western philosophy, for example, he takes as his starting point not the central arguments but rather apparently peripheral notions implied by the texts. The marginal becomes the main line and in doing so undermines the force of the original argument by showing that somehow its conclusions were already inherent in the presuppositions made (which were concealed but made apparent by the process of deconstruction). It is Derrida's argument that the texts *already always* contain their own deconstruction. One of his metaphors for deconstruction is that of a computer virus.

Foucault's work has perhaps the least *semiological* focus of poststructuralist thinkers, and yet his fascination with *discourse* is central. Foucault, in a variety of quite different (and highly original) ways, works around the relationship between power and knowledge. He actually hyphenates the two—*power-knowledge*—to emphasize their inextricable interconnection. All disciplines of knowledge have institutional manifestations of power relations, and of course, these institutions—the psychiatric clinic and the prison are two of the institutions he analyzed—are also based upon discursive structures of knowledge.

Garry Potter

See also Contemporary French Philosophy and the Social Sciences; Discourse Analysis; Durkheim's Philosophy of Social Science; Foucault's Thought; Language and Society; Semantics and Pragmatics; Sociolinguistics; Social Anthropology; Symbolism

Further Readings

Althusser, L. (1969). *For Marx* (B. Brewster, Trans.). London, England: Allen Lane/Penguin.

Barthes, R. (1968). *Elements of semiology.* New York, NY: Hill & Wang.

Derrida, J. (1978). *Writing and difference* (B. Alan, Trans.). Chicago, IL: University of Chicago Press.

Foucault, M. (1978). *Discipline and punish: The birth of the prison* (A. Sheridan, Trans.). New York, NY: Pantheon Books.

Jakobson, R. (1937). *Lectures on sound and meaning.* Cambridge: MIT Press.

Lacan, J. (1968). *The language of the self: The function of language in psychoanalysis*. Baltimore, MD: Johns Hopkins University Press.

Lévi-Strauss, C. (1966). *The savage mind.* Chicago, IL: University of Chicago Press.

Lévi-Strauss, C. (1969). *The elementary structures of kinship*. Boston, MA: Beacon Press.

Saussure, F. de (1966). *Course in general linguistics* (W. Baskin, Trans.). New York, NY: McGraw-Hill.

SUPERVENIENCE

Supervenience, as a philosophical term of art, describes a relation between sets of properties. The central idea is that properties belonging to a set *A* supervene on properties belonging to a set *B*, the so-called supervenience base, iff any two individuals or possible worlds that are exactly alike with respect to *B*-properties are exactly alike with respect

to *A*-properties. In other words, any two things that differ with respect to *A*-properties must differ with respect to *B*-properties. Mental properties can be said to supervene on physical properties, for instance, iff any two individuals who instantiate exactly the same physical properties instantiate the same mental properties.

The supervenience relation in its contemporary use first rose into prominence in metaethics. According to the British moral philosopher Richard M. Hare, any two persons or actions satisfying exactly the same nonmoral description have to be ascribed the same moral attributes; otherwise, the logical rules for the use of terms like *good* would be violated. Thus, supervenience is employed to formulate adequacy conditions on the use of moral terms. Later, Donald Davidson introduced the notion into the philosophy of mind to describe the controversial relationship between the mental and the physical. Nowadays, philosophy of mind still is one of the most important fields of application of the supervenience relation. However, the notion also plays a central role in various other fields of analytic philosophy. In contemporary metaphysics, there are debates on whether modal properties supervene on nonmodal properties and whether dispositions supervene on categorical properties. In ethics and aesthetics, supervenience is used to describe the relationship between normative and descriptive properties as well as between aesthetic and natural properties.

Definitions of Supervenience

Weak and Strong Supervenience

The intuitive definition of supervenience given above leaves it open whether the supervenience relation relies on comparisons between single individuals or between whole possible worlds. Depending on how this issue is decided, one can spell out the definition of supervenience in different ways. If supervenience is taken to rely on comparisons between single individuals, the two standard notions are weak supervenience (WS) and strong supervenience (SS), usually defined in the following way:

> *Weak supervenience:* *A*-properties weakly supervene on *B*-properties iff for any possible world w and any individuals x and y, if x and y have the same *B*-properties in w, then x and y have the same *A*-properties in w.

> *Strong supervenience:* *A*-properties strongly supervene on *B*-properties iff for any possible worlds w_1 and w_2 and any individuals x and y, if x in w_1 has the same *B*-properties as y in w_2, then x in w_1 also has the same *A*-properties as y in w_2.

To illustrate the intuition behind WS and SS, suppose that mental properties weakly supervene on physical properties and that x instantiates the mental property of *being happy* in some possible world w. Then, it follows from WS that all individuals who inhabit w and have the same physical properties as x are happy, too. The crucial difference between WS and SS is that the former only requires that individuals inhabiting the same possible world and having the same *B*-properties are alike with respect to *A*-properties, whereas SS involves cross-world comparisons between individuals. Accordingly, mental properties strongly supervene on physical properties iff any two individuals, whether they are world-mates or inhabit different possible worlds, have the same mental properties if they are alike in any physical respect. SS hence implies WS—that is, whenever *A* strongly supervenes on *B*, then *A* weakly supervenes on *B*, but the converse does not hold.

Global Supervenience

In contrast to WS and SS, which rely on comparisons between single individuals, the notion of global supervenience takes into account the distribution of properties over whole possible worlds. The underlying idea is that *A* globally supervenes on *B* iff any possible worlds w_1 and w_2 that are exactly alike with respect to *B* are also exactly alike with respect to *A*. Suppose that Φ is a set containing properties and relations of a certain type, for example, physical properties and relations. Then, the claim that w_1 and w_2 are exactly alike with respect to Φ is typically interpreted as the claim that there is a Φ-preserving isomorphism between w_1 and w_2, that is, a one–one mapping, Γ, of the inhabitants of w_1 onto the inhabitants of w_2, such that for any Φ-property F, an individual x has F in w_1 iff $\Gamma(x)$ has F in w_2 and for any Φ-relation R, the tuple $\langle x_1, \ldots, x_n \rangle$ instantiates R in w_1 iff $\langle \Gamma(x_1), \ldots, \Gamma(x_n) \rangle$ instantiates R in w_2. Accordingly, w_1 and w_2 are alike with respect to B iff there is a *B*-preserving isomorphism between them. The three

standard definitions of global supervenience all rely on this notion:

> *Weak global supervenience (WGS):* *A*-properties weakly globally supervene on *B*-properties iff for any possible worlds w_1 and w_2, if there is a *B*-preserving isomorphism between w_1 and w_2, then there is an *A*-preserving isomorphism between w_1 and w_2.
>
> *Intermediate global supervenience (IGS):* *A*-properties intermediately globally supervene on *B*-properties iff for any possible worlds w_1 and w_2, if there is a *B*-preserving isomorphism between w_1 and w_2, then there is a *B*-preserving isomorphism between w_1 and w_2 that is also an *A*-preserving isomorphism.
>
> *Strong global supervenience (SGS):* *A*-properties strongly globally supervene on *B*-properties iff for any possible worlds w_1 and w_2, any *B*-preserving isomorphism between w_1 and w_2 is an *A*-preserving isomorphism.

The notions of global supervenience are all logically weaker than the notion of SS. This becomes evident, for instance, if it is assumed that externalism about mental content is true. According to externalism (also called "social anti-individualism" and dealt with in a separate entry in this encyclopedia), mental states, such as beliefs (and other propositional attitudes entertained by a speaker/thinker), do not solely depend on an individual's intrinsic physical states (and their properties), but they also depend on the environment in which the individual is placed, that is, on the kinds of concepts available in the linguistic or conceptual environment in which the speaker finds herself (i.e., concepts or linguistic terms that appear in the propositional attitudes in question). In other words, if externalism is true, mental states (and their properties) fail to strongly supervene on intrinsic physical states (and their properties). To see this, suppose that John living on Earth and Twin-John living on Twin-Earth instantiate exactly the same intrinsic physical properties. If then John and Twin-John both believe that the transparent liquid called "water" in their community quenches thirst, John has a belief about H_2O, whereas Twin-John has a belief about XYZ. According to externalism, this means that John and Twin-John are in different belief states and consequently have different mental properties. Thus (leaving aside the fact that John's body partially consists of H_2O, whereas Twin-John's body partially consists of XYZ), John and Twin-John have exactly the same intrinsic physical properties but different mental properties, and so SS fails. The criterion of global supervenience is not violated, however. Earth and Twin-Earth, which can be considered as two different possible worlds, are radically different in some physical respect, namely, the chemical structure of the transparent liquid called "water." Therefore, there is no isomorphism preserving physical properties between them, and none of the three criteria of global supervenience will be violated if John and Twin-John have different mental properties.

In general, as long as there is no formal restriction on the structure of the properties contained in *A* and *B*, the three notions of global supervenience are logically distinct from WS and SS. Karen Bennett shows, however, that each of the definitions of global supervenience is equivalent to SS if *A* and *B* contain intrinsic properties only. Moreover, Robert Stalnaker proposes a proof to the effect that if the supervenience base contains properties involving complete *B* descriptions of possible worlds, SS is equivalent to SGS. Yet both Bennett's and Stalnaker's arguments are restricted to the case where *A* and *B* contain properties only. Oron Shagrir argues that if *A* and *B* contain relations, SGS is crucially distinct from SS, since SS is defined for properties only and SGS can make claims about dependence relationships between sets of relations, which cannot be made using SS.

Supervenience and Other Relations

Supervenience is a term of art that does not correspond to any commonsense notion. In particular, it cannot be equated with a causal or explanatory relationship. That supervenience cannot be equated with a causal relationship follows directly from the observation that supervenience is a synchronous relation whereas causality is a diachronic relationship. That supervenience claims alone are not sufficient for grounding an explanatory relation becomes clear in view of the fact that supervenience is a reflexive relation—each set of properties trivially supervenes on itself according to all the notions of supervenience specified above. However, identity does not count as an explanatory relationship. If Peter has the mental property of *being happy,* for instance, this cannot suitably be explained by the fact that Peter is happy.

Given that supervenience is not a causal or explanatory relationship, many authors claim that supervenience expresses a dependence relation: If *A* supervenes on *B,* then the properties contained in *A* are dependent on the properties contained in *B*. It is questionable, however, whether supervenience is strong enough to ground a dependence relation. If dependence is understood as an ontological priority relation, that is, in such a way that *A*'s supervenience on *B* implies that the *B*-properties are ontologically prior to or more fundamental than the *A*-properties, supervenience cannot be equated with dependence. Suppose, for instance, that *A* = {*being a father*} and *B* = {*being male, having at least one child*}. Then, *A* supervenes on *B,* and vice versa. But the mutual supervenience relation holding between *A* and *B* does not imply that the properties contained in *B* are more fundamental than the property of *being a father* or that the latter property is more fundamental than the properties contained in *B*.

It is possible, however, to interpret supervenience as a certain kind of functional dependence. Consider a complete physical description of an individual (or a possible world). Supervenience of the mental on the physical then implies that there is exactly one mental description such that any individual (or possible world) satisfying the physical description under consideration also satisfies this particular mental description. Accordingly, the supervenience relation between the mental and the physical can be interpreted as a function from combinations of physical properties to combinations of mental properties: Any possible combination of physical properties instantiated by an individual or possible world will be mapped onto exactly one combination of mental properties.

Applications

Physicalism

One major field of application of supervenience is the definition of physicalism in the philosophy of mind. Physicalism is usually understood as the thesis that fixing all the physical facts is sufficient for fixing *all* the facts. Thus, physicalists claim that the nonphysical, particularly the mental, is necessitated by and systematically covaries with the physical. This implies, for instance, that zombies, creatures that are physically like us but have no mental properties whatsoever, are metaphysically impossible. On the basis of this intuition, several authors, notably Frank Jackson and David Lewis, propose a definition of physicalism according to which physicalism is true in a world *w* iff any physical duplicate of the actual world is a duplicate *simpliciter* of *w.* This condition can be interpreted as a global supervenience claim: Physicalism is true in *w* iff each possible world that is like *w* in all physical respects is also like the actual world in all nonphysical respects. Thus, for example, if physicalism is true in the actual world, then there are no metaphysically possible worlds that are like the actual world but in which the physical duplicates of some humans are zombies. According to this definition of physicalism, humans and their physical duplicates are also alike in all nonphysical respects.

The intuitive definition of physicalism leaves open how the underlying global supervenience should be interpreted. It is widely agreed, however, that WGS and IGS are too weak to ground physicalism since they are compatible with what Karen Bennett calls *intraworld variation,* that is, two individuals inhabiting the same possible world and having exactly the same physical properties but different mental properties. Yet if there can be two individuals inhabiting the same possible world who are in exactly the same bodily state but have radically different mental properties, such that one of them feels happy, say, whereas the other feels sad, the requirement that mental properties systematically covary with physical properties is violated. The notion of SGS excludes intraworld variation. Accordingly, if physicalism is to be understood as a global supervenience claim, then global supervenience should be interpreted as SGS.

It is controversial, however, whether physicalism can adequately be defined in terms of supervenience. We will mention here two challenges. One is that a global supervenience requirement might be too strong. It captures the intuition that zombies are not metaphysically possible. However, it rules out that there are worlds that are very similar to the actual world but contain extra entities, such as angels or ghosts. Yet physicalists do not have to rule out nonphysical entities, for example, angels or ghosts, as logically or metaphysically possible. The central tenet of physicalism is restricted to the thesis that in worlds like our world, all nonphysical entities are necessitated by and systematically covary with the physical ones and therefore our world is not inhabited by any angels or ghosts. Jackson therefore proposes to restrict the supervenience thesis to worlds that are *minimal* physical duplicates of *w,* that is, worlds that are like *w* in all physical respects

and that do not contain any extra entities (a similar approach was proposed by David Lewis, Terence Horgan, and Chalmers). If only such worlds are taken into account, the definition of physicalism is compatible with the assumption that angels or ghosts are metaphysically possible.

Another challenge is that a supervenience-based criterion of physicalism might be too weak. One problem is that global supervenience theses cannot rule out the existence of necessary nonphysical beings, for example, the assumption that God exists in every metaphysically possible world, which is incompatible with physicalism. Moreover, it can be argued not only that physicalism requires the existence of an ontological covariance relation between nonphysical and physical properties but also that this covariance can be explained in a physicalistically acceptable way. Horgan therefore proposes to replace the ordinary notion of supervenience, which cannot ground an explanatory relationship, by the stronger notion of superdupervenience—a supervenience relation that does not just hold for metaphysical reasons but is in addition explainable in a physicalistically acceptable way.

However, even though it is controversial whether supervenience alone is sufficient for defining physicalism, it is widely agreed that some supervenience requirement is a necessary condition of physicalism. Even proponents of opposing views take it for granted that if there is a systematic relationship between mental and physical properties at all, then it must obey some supervenience principle. This is the reason why supervenience has a crucial role to play in the debate on physicalism and the mind–body problem.

Coincidentalism

Another important application of supervenience pertains to the debate over coincident entities. Coincidentalism is the view that there are pairs of numerically distinct entities that are made of the same parts. Coincidentalists argue, for example, that a statue and the lump of matter from which it is made are coincident yet numerically distinct entities, since they differ in their modal properties: The lump but not the statue can survive being flattened. A pattern of supervenience argument against coincidentalism is that their modal properties cannot differ. The modal difference between the statue and the lump must be grounded in qualitative nonmodal properties and relations (call it BASE), such as their subatomic structure. Modal properties, in other words, supervene on nonmodal, BASE, properties and relations. But given that the statue and the lump—as coincident entities—share exactly the same BASE at the time they coincide, we must conclude that the statue and the lump cannot differ modally.

Coincidentalists can reply that there are understandings of supervenience under which the alleged conclusion does not follow: The statue and the lump can still differ in their modal properties. Assume, for example, that modal properties weakly globally supervene on BASE. Weak global supervenience requires that BASE-indiscernible worlds be modal indiscernible. It requires that in a BASE-indiscernible world, one object has the modal properties of the statue and another the modal properties of the lump. But this requirement is certainly in accord with the entities' coincidence, as the statue and the lump can have different modal properties.

In responding, some have pointed out that not every notion of supervenience counts as an adequate notion of dependence. We must be sure, for example, that weak global supervenience relations reflect adequate grounding relations. However, Bennett's intraworld variation example (above), among other examples, gives us reason to think that weak global supervenience is not a notion of dependence and thus cannot establish the grounding of the modal in the nonmodal. More recently, Theodore Sider suggests that there is a fundamental modal *relation* between the statue and the lump of matter. This relation, of *opposite-possibly surviving being squashed*, means exactly that one might have survived being squashed. It strongly globally supervenes on BASE. The strong global supervenience of the modal (monadic) properties (*surviving being squashed*) on BASE fails. But these monadic properties are artifacts of the language we ordinarily speak, and so the failure of supervenience does not indicate a real metaphysical nondependence. Here too, supervenience does not settle the metaphysical debate, but it advances our understanding of unsettled issues.

Oron Shagrir and Vera Hoffmann-Kolss

See also Causation, Philosophical Views of; Distributed Cognition and Extended-Mind Theory; Emergence; Events; Explanation, Theories of; Metaphysics and Science; Mind–Body Relation; Reduction and the Unity of Science; Social Anti-Individualism and the Mental

Further Readings

Bennett, K. (2004). Global supervenience and dependence. *Philosophy and Phenomenological Research, 68*(3), 501–529.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory.* New York, NY: Oxford University Press.

Davidson, D. (1980). Mental events. In D. Davidson (Ed.), *Essays on actions and events* (pp. 207–225). Oxford, England: Clarendon Press.

Hoffmann, V., & Newen, A. (2007). Supervenience of extrinsic properties. *Erkenntnis, 67*(2), 305–319.

Horgan, T. (1993). From supervenience to superdupervenience: Meeting the demands of a material world. *Mind, 102*(408), 555–586.

Jackson, F. (1998). *From metaphysics to ethics: A defence of conceptual analysis.* Oxford, England: Clarendon Press.

Kim, J. (1993). *Supervenience and mind.* Cambridge, England: Cambridge University Press.

Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy, 61*(4), 343–377.

McLaughlin, B. P. (1997). Supervenience, vagueness, and determination. *Philosophical Perspectives, 11,* 209–230.

Moyer, M. (2008). Weak and global supervenience are strong. *Philosophical Studies, 138*(1), 125–150.

Paull, R. C., & Sider, T. R. (1992). In defense of global supervenience. *Philosophy and Phenomenological Research, 52*(4), 833–854.

Shagrir, O. (2002). Global supervenience, coincident entities and anti-individualism. *Philosophical Studies, 109*(2), 171–196.

Sider, T. (1999). Global supervenience and identity across times and worlds. *Philosophy and Phenomenological Research, 59*(4), 913–937.

Stalnaker, R. (1996). Varieties of supervenience. *Philosophical Perspectives, 10,* 221–241.

Symbolic Interactionism

Symbolic interactionism is a major theoretical tradition in sociology. The term was coined by the American sociologist Herbert Blumer (1900–1987) in a 1937 essay on debates in social psychology: He later described it as a "barbarous neologism" that stuck. It may be understood as the product of the encounter between American philosophical pragmatism, sociology, and social activism. This entry traces the development of symbolic interaction, the history of its older intellectual precursors, and more recent developments.

Social Behaviorism

Blumer and his closest followers identified their principal philosophical inspiration as that of George Herbert Mead (1863–1931), a close associate of John Dewey, who joined the University of Chicago's philosophy department in 1894 and remained there until his death in 1931. Mead's own preferred description for his work was *social behaviorism,* contrasted with the radical behaviorist psychology of his Chicago contemporary John B. Watson. Watson argued that "mind" was a redundant metaphysical notion. All action could be explained in terms of responses to stimuli, without the need to posit some intervening process. Mead objected that there was an important difference in the processing of incoming stimuli by humans, compared with the animals studied by behaviorist psychologists. Animals treated stimuli as *signs,* information about the world that directly triggered behavioral responses. Humans mostly responded to stimuli as *symbols,* signals that required cognitive transformation before becoming the basis of action. Mead also stressed the extent to which the meaning of acts was *not* derived from the actor's intention. An actor would design an act in *imagined* anticipation of the response of its projected recipient or audience. However, the meaning of that act would only emerge from the *actual* response and might be further revised by the producer's subsequent self-correction. The act of speech, for example, involves organizing the range of sounds producible by the human body into blocks that stand for, or *symbolize,* the actor's imaginative projection of a hearer's response. For these blocks to function as potentially shareable communicative symbols, they must form part of an *intersubjective* cultural system: Words and language would be the most familiar example. I find out who I am and what I meant from your response: This involves the so-called looking glass self, in the terms of an earlier American pragmatist, the sociologist Charles Horton Cooley (1864–1929). Each of us would ultimately be constrained by our understanding of what an idealized citizen, the *generalized other,* would be doing through the use of this symbol, given its place within our shared system. This system is a social phenomenon, something that we learn in the course of socialization.

(In this respect, Mead has an extensive discussion of children's games and their contribution to the development of skills necessary for successful interaction.)

Similar arguments were used by sociologists such as Ellsworth Faris and Luther Bernard against those who argued that all human behavior could be reduced to inherited biological drives, an early-20th-century version of genetic determinism and evolutionary psychology. Although these writers acknowledged the importance of understanding human embodiment as a material constraint on action, they argued strongly against any suggestion that biology could be preeminent.

The Wider Tradition

The emphasis on pragmatism does not do full justice to the philosophical traditions on which symbolic interactionism is based. The citation conventions of 19th- and early-20th-century academic writing have obscured the influence of ideas from Scottish Enlightenment writers, especially Adam Smith (1723–1790). Key concepts like the *looking glass self* and the *generalized other* derive from Smith's *Theory of Moral Sentiments*. This connection locates symbolic interactionism in a lineage to the ancient Stoic philosophers and also draws our attention to another line of descent from the Scottish Enlightenment: through Kant to Austrian philosophers, like Alfred Schütz and Ludwig Wittgenstein, and economists, like Friedrich Hayek and Ludwig von Mises, with their interest in language, in the nature and distribution of knowledge, and in the communication of information. This line results ultimately in symbolic interaction's sociological cousin, *ethnomethodology*.

Recent Developments

Since the late 1980s, the unity of symbolic interactionism has split along three axes: ontological; epistemological; and political. One group has tried to identify symbolic interactionism with a postmodernist version of social constructionism as a basis for critique and action. In response, many "realists" have discarded the label, while continuing to draw on its tradition in their studies of social organization. They accept Mead's analysis of the constraints created by the material world, even if this is knowable only through cultural symbols, and of the importance of true representation as a regulating ideal for inquiry, even if practically unachievable. An overlapping division is between those who accept the ethnomethodological argument that, since we cannot know another person's mind, we should not speculate about this and should only study his or her practices, and those who retain the view that empathic understanding allows us to induce the meaning of actions for participants. This is often seen in disputes over the legacy of Erving Goffman (1922–1982), whose work has influenced both symbolic interactionists and ethnomethodologists. Finally, there is division over symbolic interactionism's politics. The original Chicago work was closely associated with the social activism of Jane Addams (1860–1935) but did not follow her radical turn in the late 1920s. Contemporary symbolic interactionists have often been associated with "standpoint" research, using empathic analysis to recover the moral and cultural integrity of socially marginal or excluded groups and to challenge their "outsider" status. This sits uncomfortably with the degree to which symbolic interactionism can also be seen as part of the "spontaneous order" tradition in political thought.

Robert Dingwall

See also Behaviorism in Psychological Explanation; Ethnomethodology; Hayek and the "Use of Knowledge in Society"; Philosophy of Sociology, History of; Pragmatism and the Social Sciences; Scottish Enlightenment: Influence on the Social Sciences; Social Constructivism; Spontaneous Order

Further Readings

Berry, C. J. (1997). *Social theory of the Scottish Enlightenment*. Edinburgh, Scotland: Edinburgh University Press.

Blumer, H. (1969). *Symbolic interaction: Perspective and method*. Englewood Cliffs, NJ: Prentice Hall.

Deegan, M. J. (1988). *Jane Addams and the men of the Chicago school, 1892–1918*. New Brunswick, NJ: Transaction.

Reynolds, L. T., & Herman-Kinney, N. J. (Eds.). (2003). *Handbook of symbolic interactionism*. Walnut Creek, CA: Altamira.

Rock, P. (1979). *The making of symbolic interactionism*. London, England: Macmillan.

Rose, A. T. (Ed.). (1962). *Human behavior and social processes: An interactionist approach*. London, England: Routledge & Kegan Paul.

SYMBOLISM

This entry presents the various meanings of *sign, symbol,* and *symbolism*; charts various approaches to them; and highlights the manner in which the social sciences have approached the use of signs and symbolism as central to their domain.

Signs, Symbols, and Semiosis

Ways of perceiving symbolism range from a concept so broad that it encompasses the entire human world of meanings to one that narrows the term solely to representations of the transcendent sphere. If, thus, the word *symbol* refers to every element endowed with a meaning, then the entire sociocultural human reality may be considered in terms of symbolism.

The terms *sign* and *symbol* are often used interchangeably in the study of *semiosis,* that is, the formation of meanings. The object, fact, or event that functions as a sign or symbol refers to something other than itself. Contemporary *semiotics* or *semiology* (from the Greek σήμα), stemming from the work of Charles S. Peirce and Ferdinand de Saussure, has its roots in the ancient and medieval concepts of *signata* (from Latin *signum*—"sign") and *sumbolon* (from the Greek *σύμβολον* and the verb *sumballein*—"to put together").

The typology of signs proposed by Charles S. Peirce, who distinguished indexes, icons, and symbols understood as conventional signs, has become widespread in the theory of *semiosis* that relates to the use of signs. Semiosis implies a continuous interpretation of signs in communication. The sign in the process of semiosis is based upon its triadic structure of *representamen, object,* and *interpretant*. For Peirce, no sign is exclusively a symbol, an icon, or an index, but every sign contains—in various proportions—elements of them all.

Sign and symbol are sometimes differentiated on the ground that a sign indicates something, whereas a symbol stands for, or takes the place of, another thing. A trail marker, for example, points in a direction, and money represents a value. The emphasis in the case of the former term is put on that which is pointed out and signified and in the case of the latter is placed on the signifier, which functions as a substitute for something.

Natural language is the primary or most basic form of symbolic system. Words can be considered as linguistic signs or symbols. The common function of signs and symbols is invoking something else to come in mind. The nature of *meaning* is here the central point. A broad range of views may be categorized as either cognitive or pragmatic standpoints focusing on thought and action, respectively. The modes of signification or symbolization vary depending on the kinds of things functioning as signifiers, including discursive symbolism, symbolism of objects, and behavioral symbolism.

The symbolic functioning of things as representations has been analyzed in terms of *natural* and *conventional* signs or symbols.

Natural Signs

Natural signs enable inferences about the cause (e.g., fire) on the basis of the effect (e.g., smoke). Historically, natural signs or natural meanings were taken as the foundation of the premise about the symbolism of the whole natural reality conceived of as God's Book of Nature expressing meanings. This view has reappeared as the Romantic understanding of symbols epitomized by Friedrich Schlegel, with hieroglyphs carrying the sense of a higher reality, and likewise by G. W. F. Hegel, in his *Aesthetics,* for whom the Sphinx was a symbol of symbolization, imparting a mysterious meaning. This ontological or metaphysical understanding of symbols can also be found among the American transcendentalists, as well as the French symbolists. Of the former, Ralph Waldo Emerson stated that we are symbols and inhabit symbols, while of the latter, Charles Baudelaire used, in his *The Flowers of Evil,* the famous metaphor of the human journey through the forest of symbols.

The romantic conception of symbolism permeates the humanities, as exemplified by hermeneutics (Paul Ricoeur), religious studies (Mircea Eliade), history (Eric Voegelin), and even semiotics (Umberto Eco). A similar stance has been taken by the social phenomenologist Alfred Schütz, who placed symbols in provinces of meaning that transcend everyday life. Following this approach, the sociologists Peter Berger and Thomas Luckmann adopted a narrow conception of symbolism in the form of symbolic universes—in other words,

systems of meanings that link symbolic reality with transcendent reality.

Conventional Symbols

The *arbitrary* relationship between a symbol's meaning and the symbolized thing consists in the fact that the meaning is not contained in the symbol itself but in the symbolized thing. Furthermore, the symbol has no causal connection with its meaning and, thus, with the symbolized thing: The relationship is purely conventional.

Unlike natural signs, conventional signs or symbols are intentionally devised by human beings to signify. Linguistic symbols are particularly important because other modes of meaningful expression may be converted into words but not vice versa. Words may signify real things or name them, as well as signify ideas by expressing their meanings. Ideas as mental images are meanings and were treated as symbols of things or thoughts. Multiplicity of meanings is characteristic of linguistic symbols and allows for the ambiguous or equivocal use of words, particularly the invention of metaphors. Signs, on the contrary, are unequivocal (e.g., a red traffic light for "Stop").

Literal and metaphorical meanings provide a ground for arguments to narrow the notion of symbolism to the latter. Such an approach has its roots in the interpretation of religious sacred texts in a multiplicity of senses beyond the literal or historical, considered as spiritual meanings. The symbolic mode operates through a pragmatic decision to interpret the text symbolically beyond the literal level.

A narrower definition of the term *symbol* relies on the differentiation between two types of meaning: *denotation* and *representation*. A person's given name denotes (refers to him or her) but does not represent; in contrast, a national emblem represents the state but does not name (denote) it. Symbols represent that which they substitute for. Symbols thus defined may differ in arbitrariness. Words are generally considered wholly arbitrary symbols of ideas. Words do not denote ideas but represent them. Both words and the ideas they represent denote reality. In other words, individual terms (names) in language denote, whereas symbols represent.

Symbolism is frequently linked exclusively with complex and unobvious meanings; that which is symbolic may be contrasted with that which is real, since a symbol can substitute for something that doesn't really exist. Different conceptions of symbolism stem from research on mythology, on the one hand, and from the theory of language, on the other. Expressive symbolism conveys a meaningful emotional experience. Its antithesis is the symbolism of the completely abstract categories of scientific language. Relating to these cognitive and emotional dimensions, Edward Sapir distinguished between referential symbols and symbols of condensation. Ernst Cassirer, in his theory of symbolic forms of culture, such as language, myth and religion, art, history, and science, distinguished three types of symbolism: expressive, conceptual, and representative. The third one, characteristic of social phenomena, combines and balances the functions of the two previous ones.

The Social Sciences

Symbolic Action and Symbolic Interaction

Ways of understanding symbolism and defining symbols reflect the diversity of views about the nature of reality and cognition. The differences also stem from specific features of the studied realms of sociocultural reality, on the grounds of which conceptions of symbolism develop. Symbols and symbolic actions were especially closely studied by social and cultural anthropologists, in the field of rituals and religious myths, notably in symbolic anthropology, where symbols are not purely conceptual but have a dramaturgical and performative character, as shown by Victor Turner. Clifford Geertz says that symbols represent models of reality and models for patterned acts. Ritual remains the prototype of symbolic action, that is, of the symbolic performance, which in its elementary form is an action dealing with something that serves as a symbol and represents an analogous action dealing with the symbolized thing.

Studies on the social function of symbolization were initiated by social anthropologists such as Alfred Radcliffe-Brown and Bronisław Malinowski, as well as by cultural sociologists like Florian Znaniecki, Pitirim A. Sorokin, and Robert M. MacIver, and afterward by symbolic interactionists, who followed the American pragmatists. Among the latter, George H. Mead's notion of the conversation of gestures and significant symbols has been

particularly seminal. According to this approach, symbolism is examined in the context of coordinated actions. Symbols make it possible to anticipate the future course of interaction.

Social Symbolism and Collective Action

The analysis of symbolism's functions extends beyond linguistic theories focused on the cognitive function of language. Referring to the significance of symbols for social relations, Raymond Firth distinguished between expression, communication, knowledge, and control. The 18th-century Scottish moral philosophers Adam Ferguson and Adam Smith argued that symbols indicate social relations, consolidate social order, and represent power and wealth. It was Émile Durkheim who most persuasively claimed that social life in all its aspects is possible only due to a vast symbolism. Researchers from the Durkheim school focused not only on functions but also, and to a greater degree, on the identification of symbolic forms. Albert Salomon showed that Durkheimian collective representations include three forms: symbolic collective representing, symbolic affecting, and symbolic recollecting (collective memory).

Pierre Bourdieu developed a theory of the social symbolic system that is organized according to the logic of difference, beginning with basic categories such as gender, social hierarchy, or fellow men and strangers. It presents the problems of symbolic power, symbolic violence, and symbolic conflicts.

When one assumes the social character of language, symbolism is not contained in an autonomous, semiotical-linguistic system but is socially produced. The main focus of attention is not on social knowledge but on systems of action. The functions of symbolization in social practices are constitutive and transformative of social order. Symbolic boundaries are demarcated, abolished, or transcended.

Symbolization is not limited to writing and speaking but is contextualized in action and in acts of communication. This entails investigating symbolic practices constructing the framework of collective action and social movements—that is, the symbolic constitution of the objective of such actions and of the identity of collective agents. Power is symbolic and is accompanied by a symbolism of power. Methods of action are a variation of basic forms such as coercion and the influence through which power is manifested. Both coercion and persuasion require symbolization as a tool of influence. Power, in the sense of influencing, dominating, manipulating, or transforming collective life, remains a central problem for the sociology of symbolic processes.

Meanings are products of social acts as signifying acts and generated through interacting. Symbols are studied as instruments or tools of action where their use is decisive for the reproduction or transformation of social relations and involves strategies or politics of symbolization. Social symbolism has three basic functions—constitutive, conservative, and transformative—with respect to individuals and to collectivities.

Symbolism belongs to the anthropological order and constitutes the human sociocultural reality. It is a research subject in many of the cultural sciences, which uncover the symbolic constitution of culture. Studies of the functioning of symbols have proven significant for the basic problem of liberty, both in the private and in the public sphere.

Elżbieta Hałas

See also Communicative Action Theory; Cultural Studies; Language, Philosophy of; Metaphor; Power; Pragmatism; Semantics and Pragmatics; Social Anthropology; Social Construction of Reality; Structuralism and Poststructuralism; Symbolic Interactionism

Further Readings

Bourdieu, P. (1991). *Language and symbolic power.* Cambridge, England: Polity Press.

Cassirer, E. (1953–1957). *The philosophy of symbolic forms* (3 vols.). New Haven, CT: Yale University Press.

Eco, U. (1984). *Semiotics and the philosophy of language.* London, England: Macmillan.

Eliade, M. (1959). *Methodological remarks on the study of religious symbolism.* In M. Eliade & J. M. Kitagava (Eds.), *The history of religions* (pp. 86–107). Chicago, IL: University of Chicago Press.

Firth, R. (1973). *Symbols: Public and private.* London, England: George Allen & Unwin.

Geertz, C. (1973). *The interpretation of cultures.* New York, NY: Basic Books.

Malinowski, B. (1972). The problem of meaning in primitive languages. In C. Ogden & I. Richards (Eds.), *The meaning of meaning* (pp. 296–336). London, England: Routledge & Kegan Paul.

Mead, G. H. (1934). *Mind, self and society.* Chicago, IL: University of Chicago Press.

Peirce, C. S. (1960). *Collected papers: Vol. II. Elements of logic* (C. Hartshorne & P. Weiss, Eds.). Cambridge, MA: Harvard University Press.
Radcliffe-Brown, A. (1952). *Structure and function in primitive society.* London, England: Cohen & West.
Ricoeur, P. (1969). *The symbolism of evil* (E. Buchanan, Trans.). Boston, MA: Beacon Press.
Salomon, A. (1955). Symbols and images in the constitution of society. In L. Bryson, L. Finkelstein, H. Hoagland, & R. M. MacIver (Eds.), *Symbols and society* (pp. 103–133). New York, NY: Harper.
Sapir, E. (1934). Symbolism. In E. R. Seligman & A. Johnson (Eds.), *Encyclopedia of social sciences* (Vol. 14, pp. 492–495). New York, NY: Macmillan.
Saussure, F. de (1959). *Course in general linguistics* (W. Baskin, Trans). New York, NY: Philosophical Library.
Schütz, A. (1962). *Collected papers: Vol 1. The problem of social reality.* The Hague, Netherlands: Martinus Nijhoff.
Sorokin, P. A. (1947). *Society, culture and personality: Their structure and dynamics.* New York, NY: Harper.
Turner, V. (1967). *The forest of symbols: Aspects of Ndembu ritual.* Ithaca, NY: Cornell University Press.
Voegelin, E. (1987). *The new science of politics: An introduction.* Chicago, IL: University of Chicago Press.
Znaniecki, F. (1952). *Cultural sciences: Their origin and development.* Urbana: University of Illinois Press.

Systems Theory

Systems theory is a multifaceted, transdisciplinary approach to the study of complex systems. Although it has roots in earlier philosophical traditions, such as the process philosophies of G. W. F. Hegel and Alfred North Whitehead, it emerged as a distinct field in the mid 20th century. It encompasses a number of different schools of thought, and the term *systems theory* is often used interchangeably with *systems thinking* or the *systems approach*.

Systems theory grew out of a recognition of the limitations of classical science in dealing with complex systems, such as biological organisms, social systems, and the increasingly sophisticated organizational and technological systems of the 20th century, often referred to as "socio-technical systems."

Building on the work of René Descartes, classical science emphasized the analytical method, breaking things down into their component parts in order to understand the whole. Because complex systems involve highly interdependent interactions among their parts, systems theorists argued that this reductionist approach was inadequate in explaining the dynamic behavior exhibited by these systems and that such systems must be understood as whole systems, as reflected in the often-quoted maxim that "the whole is greater than the sum of its parts." Systems theory, then, provides a holistic framework for understanding the organizing relationships within and among such systems, highlighting as well the relationship between any system and the environment within which it exists.

Systems theory is an inherently interdisciplinary orientation, emphasizing the need to integrate perspectives from different fields. As a result, it has contributed significantly to the evolution of both theoretical and applied social science, in addition to providing new conceptual frameworks for philosophical inquiry. It seeks to articulate principles that are common to all types of systems and to create a framework for dialogue among scholars from different disciplines in order to overcome a perceived fragmentation in knowledge that undermines humanity's ability to address the problems confronting society.

This entry provides an overview of the evolution of systems ideas, a summary of the key concepts to emerge out of this tradition, and an examination of the implications of systems theory for philosophy and the social sciences.

A Brief History of Systems Theory

Systems theory emerged in the mid 20th century, drawing on parallel developments in the fields of biology, psychology, management, engineering, information science, and ecology, as well as on a trend toward increasing interdisciplinary collaboration in the social and behavioral sciences. The broad range of perspectives that scholars from these various fields brought to their understanding of systems resulted in an enormous variety of interpretations of systems theory, leading to multiple and often contradictory implications for philosophy and the social sciences.

Ludwig von Bertalanffy (1901–1972) is generally recognized as the father of General Systems Theory, which is one of the key strands in the threads that have been woven into the systems field. He introduced the term in a seminar at the University of Chicago in 1939 to describe an approach to studying the nature of organization in all types of systems.

A *system* can be defined as any entity that is composed of parts whose interactions form an integrated whole. It is characterized by a particular structure that constrains the relationships between the elements that constitute the system and a boundary that delineates the system in the context of its environment. Systems can be either open or closed; open systems allow for the exchange of matter, energy, and information with the environment. The elaboration of the concept of open systems was one of Bertalanffy's most important contributions to the development of systems theory.

Bertalanffy believed that all phenomena could be understood as systems and that all systems (physical, biological, and social) displayed common patterns, behaviors, and properties. At the same time, as a theoretical biologist, he also believed that there were properties unique to biological systems that could not be explained according to (or reduced to) purely physical and chemical interactions. Similarly, he argued that human systems (psychological, social, and cultural) could not be explained in purely biological terms. The potential contradiction between these two positions—that all systems contain similar properties and that biological and social systems possess unique characteristics—reflects the underlying tensions between various interpretations of the systems field. In order to understand this tension, and the implications for the application of systems ideas in philosophy and the social sciences, it is helpful to explore the properties that all types of systems share in common, the most significant of which is the idea of self-regulation through feedback.

Homeostasis and Feedback: Nonlinear or Circular Causality

Another limitation of classical science, according to systems theorists, is its emphasis on linear chains of causality. Developments in the fields of biology, neurophysiology, engineering, and information science highlighted the significance of recursion and self-referentiality in the organization and function of living organisms as well as in the newly emerging technologies. There are several different schools of thought that grew out of different ways of understanding this phenomenon, and different branches of systems theory evolved out of these various formulations.

One of the most influential concepts to come out of the study of biological systems was the notion of *homeostasis,* which is the ability of living organisms to maintain themselves in a steady state, despite constantly changing internal and external conditions. This ability depends upon a system of communication among the various organs in the body that allows it to maintain constant levels of critical factors such as blood sugar, pH balance, and temperature. If any of these factors deviates from the optimal level, the body is able to respond to restore the balance. The existence of such mechanisms implies a kind of goal-directedness or purposefulness that does not exist in closed systems and depends not only upon the exchange of matter and energy in the system but also on the exchange of information. This organismic model of feedback was central in the development of James Grier Miller's (1916–2002) formulation of systems theory in his massive tome *Living Systems Theory,* as well as in the development of Talcott Parsons's social system theory.

Miller's work grew out of a decade-long collaboration among scholars from biology, psychology, and the social sciences at the University of Chicago, inspired in part by Bertalanffy's proposal of a general theory of systems. In an effort to articulate similar processes at different levels of organization in living systems, Miller identified 20 different subsystems that processed the input, output, and processing of matter, energy, and information at the level of the cell, organ, organism, group, organization, society, and what he called the supranational system. The concept of organizational levels is another important insight—namely, that living systems, as well as complex technological systems, generally consist of nested systems, with smaller systems existing within the context of larger systems, which are themselves part of systems at an even larger scale of organization.

Cybernetics

A parallel development that influenced the community of scholars at the University of Chicago who provided the foundation for Miller's work was a series of conferences sponsored by the Macy Foundation. These conferences, held in the late 1940s and early 1950s, brought together scholars from the social sciences, neurophysiology, and systems engineering, including figures such as Gregory Bateson, Margaret Mead, John von Neumann, and Norbert Wiener. The working title of the conferences, "Feedback Mechanisms and Circular Causality in Biological

and Social Systems," evolved over the course of the conferences. At one point, the group adopted the term *teleological mechanisms,* underscoring the potential for purposive activity in biological and even technological systems. Later, they proposed the term *cybernetics* to characterize the phenomena they were exploring.

Specifically, the participants in these conferences were exploring parallels between neural networks in the brain and recursive operations in the newly emerging field of computer science. This led to a growing emphasis on the role of information and communication in complex systems. In 1929, Leo Szilard had suggested that information was distinct from matter and energy; while the latter can be neither created nor destroyed, information has the potential to increase over time, providing an explanation for the phenomena of evolution and learning in living systems. The cybernetics group sought to understand the mechanisms of information processing, exploring the ways in which information is embedded in the dynamic processes that give rise to complex patterns of organization.

As might be expected, there were significant differences between the orientations of the social scientists and the engineers. Bateson and Mead, in particular, were interested in the unique role of language in the creation of social structure in human communities. In contrast, Weiner emphasized the similarities between organisms and machines. Although his view does not accurately reflect the general orientation of the field as it evolved, his name is perhaps the one most commonly associated with the concept, and the title of his book, *Cybernetics: Or Control and Communication in the Animal and the Machine,* reflects a mechanistic and reductionist orientation that was not shared by other researchers. Many cybernetics scholars were closely affiliated with the general systems community and tended to focus more on the concept of second-order cybernetics, which emphasized self-referentiality, or the importance of including the observer in the system being observed.

System Dynamics

Somewhat distinct from the cybernetics and general systems orientations, *system dynamics* emerged in the 1950s as another approach to understanding feedback processes. Based on the work of Jay Forrester, system dynamics was rooted in the circuitry models of electrical engineering and emphasized the internal dynamics of organizations rather than the exchange of information in systems. Specifically geared toward applications in management, it tended to focus on the input, processing, and output of material in production processes and to highlight the importance of understanding stocks and flows of materials in such systems. Unlike the models of feedback based on the concept of homeostasis, which only focused on negative feedback (or deviation-minimizing feedback), system dynamics identified both negative and positive (or deviation amplifying) feedback in the dynamics of systems.

Open Systems and Emergence

While feedback processes can be found in all types of complex systems, some systems thinkers were more concerned with the unique characteristics, or emergent properties, that can be found at higher levels of organization (i.e., biological organisms, human personality structures, and social systems). Bertalanffy's conception of open systems is the foundation for the concept of *emergence,* which is central to his understanding of General Systems Theory. According to the second law of thermodynamics, the entropy of any closed system will always increase. Because it takes energy to maintain any kind of organizational structure, the infamous second law implies that all systems will tend toward greater disorder. While this seems to contradict the evidence of evolution, since life has evolved increasingly complex forms over time, Bertalanffy's insight was to suggest that living organisms are open systems and are thus able to import energy from the environment and to export their entropy (or waste), allowing them to maintain complex organizational structures and, more important, to develop increasingly complex structures.

Open systems provide a context in which qualitatively new properties can emerge from the interaction of components within the system, which cannot be predicted or explained based on understanding the components alone. This perspective allows for creative and spontaneous activity in living organisms and suggests that systems at higher levels of organization possess qualities and capacities that do not exist at lower levels. Bertalanffy contrasted his theoretical orientation, which emphasized the self-organizing nature of living systems, with Miller's

organismic model, which emphasized equilibrium models of feedback and did not allow for the possibility of change. Instead, Bertalanffy argued, living organisms exist in a dynamic steady state, with the potential to adapt to changes in the environment. Ilya Prigogine (1917–2003), the Belgian physical chemist, built on this concept of open systems, suggesting that the further systems are from equilibrium, the greater the potential for more complex forms of organization to emerge. He introduced the concept of *bifurcation,* which described the tendency, as systems become increasingly unstable, to either reorganize into more complex structures or collapse into less highly ordered structures.

Relevance in the Contemporary World

Since its origins in the mid 20th century, systems theory has influenced a broad range of disciplines. Bertalanffy distinguished between three major strands of systems thinking: systems technology, systems science, and systems philosophy. As systems technology, it has been a central framework for the emergence of information technology and systems engineering in general, which has recently launched a new field of study known as "system of systems," acknowledging the highly interdependent, nested, and networked nature of current technological and organizational systems. As systems science, it articulated a more holistic paradigm for research, influencing recent developments in systems biology, as well as chaos and complexity theories. Although his work originated in the context of theoretical biology, Bertalanffy was particularly concerned with the philosophical and social implications of systems theory and emphasized the holistic and humanistic orientation of General Systems Theory as he conceived it.

Philosophical Implications of Systems Theory

The systems perspective offers significant insights into the ontological, epistemological, and ethical dimensions of philosophical inquiry. As ontology, it suggests that phenomena must be understood in terms of whole systems, in contrast to the reductionist and mechanistic orientation of classical science. Echoing insights from the field of quantum mechanics, it emphasizes the interconnected and interdependent nature of reality, proposing a cocreative relationship between the whole and the parts and integrating both upward (from part to whole) and downward (from whole to part) causality. It is process oriented, highlighting the emergence of organization out of the dynamic patterns of relationship between the components of a system.

The various traditions of systems thinking offer a variety of interpretations of the epistemological implications. Reflecting further parallels with emerging understandings from quantum mechanics, some schools of systems thought highlight the active role of the observer in the system being observed, leading them to embrace a constructivist epistemology. Other schools of thought, such as Miller's living system orientation and the system dynamics tradition, embody a more objectivist epistemology. Nevertheless, most systems thinkers emphasize the importance of integrating multiple perspectives in understanding any system, arguing that no single lens can provide a comprehensive and accurate representation. In addition to the implications of this orientation for education, there are significant ethical consequences that inform contemporary applications of systems thinking in social organizations. Proponents of this view consider an ethic of inclusiveness and collaboration to be essential in the application of systems thinking in the social context.

Systems Theory in the Social Sciences

The concept of emergence is central in understanding the implications of systems thinking in the social sciences. It informed Bertalanffy's views on psychology, which had a significant impact on the evolution of the fields of humanistic psychology and family systems theory. He was particularly opposed to the behaviorist model of stimulus and response as the primary motive forces in human behavior, which he saw as a reductionist approach rooted in a homeostatic model. Instead, he saw human consciousness as an emergent property characterized by self-reflective awareness, highlighting the importance of considering the role of subjectivity and autonomy in understanding humans as active agents. At the same time, he rejected the individualistic conception of identity and motivation, which conceived individuals as separate from the entire web of relations in which they are embedded.

In contrast to Bertalanffy's view, some systems models, paralleling structural and functionalist schools of thought, tend to minimize the autonomy

of the individual, emphasizing instead the structure of the social system as a whole. This orientation can be traced back to Herbert Spencer's organismic understanding of the social order and is reflected to some extent in Miller's living systems model and Talcott Parsons's social systems theory, both of which were rooted in organismic models that tended to emphasize homeostasis. Parsons's theory of social action described the interrelationships between organism, personality, culture, and society, portraying society as an autonomous system with the goal of maintaining stability, order, cooperation, and consensus through the communication of values and cultural norms. As a result, it tended to downplay the autonomy of the individual, and critics argued that the model ignored the role of conflict, did not adequately account for change, and tended to reinforce the status quo.

Building on Parsons's model, Niklas Luhmann is the most well-known contemporary social systems theorist, whose work is further informed by the conceptual framework developed by Humberto Maturana and Francisco Varela. They coined the term *autopoiesis* to refer to the process of self-creation or self-production in living systems. Although echoing earlier work on the concept of self-organizing systems, they argued that the organization of a system could not be changed without destroying the integrity of the system. In their view, living systems were closed in terms of their internal organization but open in terms of their structural composition and metabolism, reproducing themselves through a process of structural coupling with their environment. Based on this understanding, Luhmann conceived of society as an autopoietic system of communication. Like Parsons's model, this conception tended to privilege the social system as a whole over the autonomy of its individual members.

In his work on the role of communication in society, Jürgen Habermas voiced concerns that many scholars shared about the eclipse of subjectivity—or what he called the *lifeworld*—in systems models, and he engaged Luhmann in an ongoing debate about the relative significance of lifeworld and system concerns. While both addressed the source of meaning and motivation in human behavior, Habermas argued that Luhmann placed too much emphasis on the structure-maintaining role of the autopoietic social structure, emphasizing instead the generation of the social world through the evolving process of discourse. Habermas's critique was particularly influential in the application of systems thinking in social organizations.

Applied Systems Theory

While systems thinking has informed the evolution of disciplines across the spectrum, it has perhaps had the most influence in the applied social sciences. Beginning in the 1940s and 1950s as systems engineering, systems analysis, or operations research, it embodied a positivistic orientation, using systems principles to maximize performance in socio-technical systems. These approaches, which were characterized by top-down decision making and control, later became known as *hard* systems methodologies, in contrast to the *soft* systems approaches that emerged in the 1970s and 1980s, which sought to integrate the experiential, subjective (or lifeworld) dimension of organizational systems.

Soft systems approaches emphasized the active inclusion of all parts of the system in the decision-making process, recognizing not only the goal-seeking but also the relationship-maintaining function of organizations and embodying a shift in emphasis from structure to process. A critical insight to emerge from the soft systems orientation, reflecting the interpretive orientations in the social sciences, is the importance of surfacing the various beliefs, assumptions, and frames of reference—or mental models—that members of organizations bring to their participation. This has resulted in the understanding of organizations as learning systems, highlighting the potential for the emergence of new forms of social organization. More recent developments in applied systems thinking, described as *critical* or *emancipatory* systems approaches, suggest that soft systems approaches have failed to address underlying issues of power, privilege, and domination, with the ultimate goal of facilitating the design of human systems that effectively serve the whole system, while acknowledging the integrity and autonomy of the individual.

Debora Hammond

See also Complexity; Complexity and the Social Sciences; Emergence; Feedback Mechanisms and Self-Regulatory Processes in the Social Sciences; Holism, in the Social Sciences; Laws Versus Teleology; Luhmann's Social Theory; Spontaneous Order; Structural Functionalism in Social Theory

Further Readings

Bateson, G. (1972). *Steps to an ecology of mind*. New York, NY: Ballantine Books.

Bausch, K. C. (2001). *The emerging consensus in social systems theory*. New York, NY: Kluwer Academic/ Plenum.

Bertalanffy, L. von. (1968). *General system theory: Foundations, development, applications*. New York, NY: George Braziller.

Habermas, J. (1985). *The theory of communicative action: Vol. 2. A critique of functionalist reason* (T. McCarthy, Trans.). Boston, MA: Beacon Press.

Hammond, D. (2003). *The science of synthesis: Exploring the social implications of general systems theory*. Boulder: University of Colorado Press.

Hammond, D. (2005). Philosophical and ethical foundations of systems thinking. *Triple C: Cognition, Communication, Co-operation, 3*(2). Retrieved from http://www.triple-c.at/index.php/tripleC/article/view/20

Luhmann, N. (2012). *Introduction to systems theory* (P. Gilgen, Trans.). Hoboken, NJ: Wiley.

Maturanam, H., & Varela, F. (1992). *The tree of knowledge: The biological roots of human understanding*. Boston, MA: Shambhala.

Miller, J. G. (1978). *Living systems*. New York, NY: McGraw-Hill.

Parsons, T. (1971). *The system of modern societies*. Englewood Cliffs, NJ: Prentice Hall.

Prigogine, I., & Stengers, I. (1984). *Order out of chaos: Man's new dialogue with nature*. New York, NY: Bantam.

Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine*. Cambridge: MIT Press.

Tacit Knowledge

The term *tacit knowledge* was coined by Michael Polanyi to describe the elements of scientific knowledge that cannot be, or normally are not, set out explicitly but that are nevertheless essential to the activity of science and indeed all other human activity. Polanyi's slogans included the statements "We know more than we can say" and "All knowledge is either tacit knowledge or rooted in tacit knowledge." The range of application of the concept and the ground for tacit knowledge are matters on which users of the term differ, but the starting point—the most common motivation—is this: Some activity, inference, or communicative act depends on both the user and the recipient possessing some skill, inferential element, or mechanism that allows them to understand, anticipate, cooperate, or coordinate with another.

The standard example of tacit knowledge is the knowledge one needs to ride a bicycle. It is essential to the activity: Some people acquire it, and others do not acquire it. We normally use questions like "Do you *know* how to ride a bicycle?" so it is not odd to call this "knowledge," however much it departs from the paradigmatic definition of knowledge as justified true belief. We cannot articulate this knowledge and therefore cannot transmit it through words, though we may be able to coach a person learning to ride a bicycle.

Tacit knowledge is itself a paradoxical term. Standard philosophical usage treats knowledge as "true belief." But tacit knowledge is neither belief, which implies something explicit or self-conscious, nor necessarily "true"—indeed it is odd to think of calling something that cannot be derived from or contradict a true statement either true or false. The term itself is an analogical term, which likens the content in the tacit background to content whose character we understand, namely, explicit belief. Yet the term has caught on, for good reasons. The paradoxical language points to the fact that this kind of "knowledge" is unlike other "beliefs" in many ways *because* it is tacit: It cannot be examined and revised in the same way, it is not open to discussion, and it is not open to justification.

The term is an attempt to designate something that is familiar and difficult to deny. Consider a simple example: The police stop me after I thread my motorcycle through traffic and ask, "Do you know how fast you were going?" I cannot tell what was on the speedometer, but I had to know how fast I was going to successfully negotiate the traffic—if I didn't know, I would have crashed. This is not only tacit but also "nonconceptual" knowledge. In this, case, what I know is not something I could articulate even if I wanted to. These cases, which might be called the "bicycle cases"—after their paradigmatic instance, knowledge of how to ride a bicycle—constitute a large class. There is a dispute about whether this class should be taken as a model for other forms of tacit knowledge. But the existence of this kind of knowledge and its external properties, for example, the fact that it is learned as a skill, with some coaching or interaction but largely through experience, is not a matter of dispute.

The oddity of the notion of tacit knowledge is illustrated in the problem of the ambiguity between

psychological and logical notions of tacit knowledge, which bears on the question of what one is doing when the tacit is articulated. On the surface, when one explains something by making it explicit, one is dipping into one's mental world to pull out something that is already there but not attended to, on the analogy of the distinction between occurrent and dispositional belief. But the bicycle-riding case does not fit into this: There, the problem is that there is no way to formulate the knowledge in words. But there is also a problem with cases of articulation.

When we explain ourselves to someone who does not understand us, are we articulating something that we think our audience will understand better, which is to say inventing a new articulation for a new audience—in short, is our statement of tacit knowledge, as in the syntactic structure case, like a theory of the content? Is the "content" relative to our idea of the audience, not something fixed that is accessed in the mind, or is it merely a kind of comment on the failure of inference on the part of the audience, with no psychological significance? Answers to this question differ. One line is that there is a psychological fact and that it is in more or less the same form as that which is articulated. Another is that there is some psychological fact, as there is with the bicycle case, but there is no close connection between the psychological fact and the articulation, which merely patches up the inference.

The role of tacit knowledge is acknowledged in a large set of literatures, notably the construction of expert systems in artificial intelligence, in the economics and social studies of science (because of the need and, therefore, the economic value of specialized tacit knowledge in science), and especially in organization studies, where the growth and sharing of tacit knowledge in an organization is seen as a valuable benefit realizable only in social networks. Ikujiro Nonaka developed the socialization, externalization, combination, internalization model for the process of one of the most widely cited theories in knowledge management to account for the interactive sharing of tacit knowledge and the growth of tacit knowledge in organizations.

Stephen Turner

See also Common Knowledge; Embodied Cognition; Knowing-How Versus Knowing-That; Nonconceptual Content; Philosophy of Expertise

Further Readings

Collins, H. (2010). *Tacit and explicit knowledge.* Chicago, IL: University of Chicago Press.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge creating company: How Japanese companies create the dynamics of innovation.* New York, NY: Oxford University Press.

Polanyi, M. (1958). *Personal knowledge: Towards a post-critical philosophy.* Chicago, IL: University of Chicago Press.

Turner, S. (1989). Tacit knowledge and the problem of computer modelling cognitive processes in science. In S. Fuller, M. de Mey, T. Shinn, & S. Woolgar (Eds.), *The cognitive turn: Sociological and psychological perspectives on science* (pp. 83–94). Dordrecht, Netherlands: Reidel.

Turner, S. (2011). Starting with tacit knowledge, ending with Durkheim? *Studies in History and Philosophy of Science, 42,* 472–476.

Team Reasoning

In Decision Theory, it is almost universally presupposed that agency is invested in individuals: Each person acts on her own preferences and beliefs. A person's preferences may take account of the effects of her actions on other people; she may, for example, be altruistic or have an aversion to inequality. Still, these are *her* preferences, and she chooses what *she* most prefers. Opposing this orthodoxy is a small body of literature that allows *teams* of individuals to count as agents and that seeks to identify distinctive modes of team reasoning, which are used by individuals as members of teams.

This entry explains the motivation for and the general principles of team reasoning. It then presents the two leading theories of team reasoning, those of Michael Bacharach and Robert Sugden, which differ in important ways regarding how group agency comes about, what happens when there is no common knowledge of group membership, and what the group agent should take as its goals.

Team Reasoning

One motivation for theories of team reasoning is that there are games of cooperation and coordination that are puzzles for orthodox Decision Theory,

in the sense that there exists some strategy that is at least arguably rational and that a substantial number of people play in real life but whose rationality Decision Theory cannot explain and whose play it cannot predict.

One such puzzle is the game of Hi-Lo. In Hi-Lo, both (*high, high*) and (*low, low*) are *Nash equilibria*; each player has achieved the best possible payoff for herself given the action of the other player (see Table 1). Intuitively, it seems obvious that each player should choose *high* because both prefer the outcome of (*high, high*) to that of (*low, low*); but that "because" has no standing in the formal theory. Standard game theory has no way of recommending or predicting one equilibrium over the other.

The source of the puzzle seems to be located in the mode of reasoning by which, in the standard theory, individuals move from preferences to decisions. In the syntax of game theory, each individual must ask separately, "What should *I* do?" In Hi-Lo, the answer to this question is indeterminate. Theories of team reasoning extend game theory to allow the players to ask, "What should *we* do?" In Hi-Lo, the answer to this question is surely "Choose (*high, high*)."

The basic idea of team reasoning is that when an individual reasons as a member of a team, she considers which *combination* of actions by members of the team would best promote the team's objective and then performs her part of that combination. It is still *instrumental practical reasoning,* where conclusions about what an agent ought to do are inferred from premises that include propositions about what the agent is seeking to achieve, but it allows that groups can be agents with group goals that provide their standards of success. When a group of people team reason, the rationality of each individual's action derives from the rationality of the joint action of the team.

Table 1 Hi-Lo

		Player 2	
		high	*low*
Player 1	*high*	2, 2	0, 0
	low	0, 0	1, 1

Source: Author.

Bacharach's Circumspect Team Reasoning

Bacharach presents the most comprehensive, formal theory of team reasoning. For Bacharach, people "team reason" when they "group identify," that is, see themselves as strongly embedded in, or identified with, a group; whether a particular player identifies with a particular group is a matter of "framing." A *frame* is the set of concepts a player uses when thinking about her situation. To team reason, a player must have the concept "we" in her frame. Bacharach proposes that the "we" frame is normally induced or *primed* by games like Hi-Lo.

Although Bacharach proposes that some games increase the probability of group identification, he does not claim that they *invariably* prime the "we" frame. The "we" frame *might* be primed, but alternatively, a player may see the game as one to be played by two separate individual agents. Bacharach models the psychology of group identification as a random process that, independently for each member of the group, determines whether or not that individual identifies with the group. Then an individual who group identifies will maximize the expected value of the group payoff function given the probabilities that other group members fail to identify. This means that a player may find ex post that she has team reasoned when the other player has not, leading to a worse outcome for herself than if she had not team reasoned—as sometimes happens in the experiments whose results Bacharach sought to explain.

Sugden's Mutually Assured Team Reasoning

On Sugden's account of *mutually assured team reasoning,* a person will not commit himself to team reasoning unless he has assurance that others will also act on team reasoning. Sugden uses a theoretical framework in which the central concept is *reason to believe.* To say that a person has reason to believe a proposition *p* is to say that *p* can be inferred from propositions that he accepts as true, using rules of inference that he accepts as valid. In mutually assured team reasoning, team members will not act on the results of team reasoning unless each has reason to believe of all the others that (a) they identify with the group and acknowledge the group payoff function as the objective of the group and (b) they endorse and act on mutually assured team reasoning. So if Sugden's group members are not sure that they will

all cooperate to achieve what they all take to be best for the group, then they will not team reason.

Because Sugden's team reasoners cooperate in a mutually advantageous enterprise, he would also constrain the group goal so that team reasoning is welfare increasing for its members, by their own individual lights.

Natalie Gold

See also Collective Agents; Collective Goals; Collective Intentionality; Collective Rationality; Common Knowledge; Cooperation/Coordination; Game-Theoretic Modeling; Group Identity; Theory of Teams; We-Mode, Tuomela's Theory of

Further Readings

Bacharach, M. (2006). *Beyond individual choice: Teams and frames in game theory* (N. Gold & R. Sugden, Eds.). Princeton, NJ: Princeton University Press.

Gold, N. (2012). Team reasoning, framing and cooperation. In S. Okaska & K. Binmore (Eds.), *Evolution and rationality: Decisions, cooperation and strategic behavior* (pp. 185–212). Cambridge, England: Cambridge University Press.

Gold, N., & Sugden, R. (2007). Collective intentions and team agency. *Journal of Philosophy, 104*(3), 109–137.

Gold, N., & Sugden, R. (2008). Theories of team agency. In F. Peter & H. B. Schmid (Eds.), *Rationality and commitment* (pp. 280–312). Oxford, England: Oxford University Press.

Sugden, R. (1993). Thinking as a team: Toward an explanation of nonselfish behavior. *Social Philosophy and Policy, 10,* 69–89.

Sugden, R. (2003). The logic of team reasoning. *Philosophical Explorations, 6,* 165–181.

Technological Convergence

This entry presents a very recent phenomenon, technological convergence, and the key epistemological and policy issues it raises. The term *technological convergence* describes the merging of different fields of technology as a result of scientific-technological progress. This entry describes the differences of vision between philosophy and nonphilosophical standpoints and highlights the importance of recent developments in the philosophy of technology and the significance of technological convergence for a novel understanding of technologically changing societies.

The currently dominant convergence story concerns the so-called NBIC convergence, referring to the synergistic combination of *N*anotechnology, *B*iotechnology, *I*nformation technology, and *C*ognitive neurosciences. Nanotechnology would provide the necessary key competence and play the role of an "enabling technology." According to this line of thought, convergence would mean providing a common ground for the scientific disciplines involved (or for science generally), so that all four elements converge in an engineering science working at the level of atoms and molecules. Far-reaching futuristic expectations are associated with this convergence, from "human enhancement" via an almost complete elimination of aging to the solution of virtually all sustainability problems.

In philosophical terms, the main idea behind this image of technological convergence can be called *atomic reductionism*. Operating with atoms and molecules thus becomes an engineering version of a "theory of everything." It would not be a theory of physical fundamentals providing explanations at an abstract level but a practical theory of operation and manipulation. Its subject would be constructing matter in a targeted fashion at the level of its building blocks—atoms and molecules, independent of whether these would be parts of crystals and metals or of living systems. In a radicalized version of 19th-century materialist reductionism, the spheres of the living and of the social are supposed in this perspective to be explained and made the object of technical manipulation starting from their atomic basis.

This concept of technological convergence not only expresses the scientific-philosophical viewpoints of its authors and exponents but also is an instrument of research policy and the subject of political dispute. The American perspective on technological convergence, which emphasizes the momentum of scientific development as a driving force, has been confronted with a European point of view that places the *design* of the convergence process for the achievement of societal goals at the center. The idea of technological convergence has also had an impact in terms of ideology. The transhumanist movement perceives the convergence hypothesis as offering a key possibility to technically improve humans from their condition as, in Arnold Gehlen's phrase, "deficient beings," first with converging technologies and eventually to overcome many human limitations

altogether in the course of the further development of a technical civilization.

An effect of the rapid and worldwide spread and discussion of controversial notions of technological convergence has been to raise interest in the field of science and technology studies (STS). Owing to technological convergence, it has been possible to observe concretely how futuristic visions have exercised real power and have practical effects—for example, for the promotion/funding of research or in structuring societal debates about the future.

In this way, visions of the future as a medium of societal conflicts, for the preparation of future-relevant decisions, and as a tool for societal self-understanding became a focal point of STS studies. Researchers have focused mainly on the emergence conditions, spread, real impacts, and strategic roles of these visions.

In contrast, the philosophical research interest is focused on questions of validity and on the contents of the visions of technological convergence. As an element of vision assessment, the hermeneutics and epistemology of these visions play a part in any reconstruction of the vision's contents and backgrounds and thereby contribute to an increase of transparency in societal debates on scientific-technological progress in the field of NBIC convergence. This enlightenment sometimes involves the formation of new and cross-sectional subspecialties of philosophy, such as neurophilosophy or nanoethics. In general, technological convergence forces the cooperation of philosophical subdisciplines that have hitherto been operating more or less separately: applied ethics, philosophy of technology, philosophy of mind, anthropology, epistemology, and hermeneutics.

The fascinating new opportunities of manipulating matter by converging technologies in abiotic and biotic systems have also motivated thinking about possible deeper changes in human civilization and its relationships to nature and technology. Technological convergence can allow disruptive increases of the agency of human beings and thereby abruptly raise the contingency in the *conditio humana*. What this may mean is the subject of philosophical debate. Roughly, three different positions have been expressed so far. First, there are assumptions that a new wave of Baconism is arising. Nanotechnology assumed that the ability of "shaping the world atom by atom" could be interpreted as a new manifestation of optimism by making everything appear to be technically possible and controllable. Second, the opposite story starts from the "enabling" character of nanotechnology and assumes creation of the greatest uncertainty imaginable: Everything could be possible, and probably nothing could be controlled. The third story regards nanotechnology as a "cipher of the future" that serves as a catalyst for societal, philosophical, and scientific debates on issues such as the future relationship between humans and technology and the future of human nature in avoiding strong substantial claims about controllability or other issues.

These debates are useful for the self-understanding of modern society regarding its transformation through converging technologies. They require philosophical reconstruction as well as socio-scientific analysis of the different forms of (mostly visionary) communication. In this way, hermeneutical reflection based on philosophical and social science methods, such as discourse analysis, can prepare the groundwork for applied ethics and for technology assessment. Ultimately, this promotes a democratic debate on scientific-technological progress by investigating alternative approaches to the future of humans and society with or without different techno-visionary developments.

Armin Grunwald

See also Artificial Intelligence; Naturalism in Social Science; Reductionism in the Social Sciences; Technoscience and Society

Further Readings

Fiedeler, U., Coenen, C., Davies, S. R., & Ferrari, A. (Eds.). (2010). *Understanding nanotechnology: Philosophy, policy and publics*. Heidelberg, Germany: AKA.

Grunwald, A. (2007). Converging technologies: Visions, increased contingencies of the conditio humana, and search for orientation. *Futures, 39,* 380–392.

Kaldis, B. (2010). Converging technologies. In D. Guston (Ed.), *Encyclopedia of nanoscience and society.* Thousand Oaks, CA: Sage.

Nordmann, A. (2004). *Converging technologies: Shaping the future of European societies*. (High Level Expert Group "Foresighting the New Technology Wave" Report). Brussels, Belgium: European Commission.

Roco, M. C., & Bainbridge, W. S. (Eds.). (2002). *Converging technologies for improving human performance*. Arlington, VA: National Science Foundation.

Technoscience and Society

Technoscience as a label highlights the fluidity of the interactive process of science and technology and the ways in which it permeates every facet of our life—from transportation and communication to health and entertainment. Technoscience can be understood historically, sociologically, and ethically. This entry reviews all three standpoints from which technoscience can be viewed.

The scientific revolutions of the 16th century, especially Francis Bacon's concern with observation and experimentation as the proper scientific method, as well as the concerns with the power that is embedded in scientific knowledge, can be seen as forerunners of the term *technoscience*. Some early-20th-century thinkers, such as Gaston Bachelard, used the term, and others, notably Karl Popper, later linked the scientific method with political views of an "open society" while retaining the difference in aims that science and technology have.

More direct uses of the term are associated first with the Belgian philosopher Gilbert Hottois and the French philosopher Jean-François Lyotard, who used the term to combine the social, political, and economic activities within which science and technology operate. The Paris school (Bruno Latour, Michel Callon, Steve Woolgar) and the Edinburgh school of the Strong Program in the Sociology of Knowledge (David Bloor, Barry Barnes) in their sociological insights into the social construction of scientific knowledge in modernity made use of the term as well. From its European origins, the term is simultaneously sociological and critical, contextualizing the acquisition of knowledge, meaning, and truth.

Historically, the standard notion has been that science is theoretical and therefore pure or that technology is simply the application of scientific theories and principles, hence practical and messy (even when philosophically the aims remain distinct). In fact, theoretical insights arise out of experimentation just as much as out of contemplation. Isaac Todhunter has documented a series of historical moments wherein great thinkers were commissioned to find a theory, principle, or equation that would explain and predict how to win at gambling, ensure financial solvency in cases of overseas transportation, and collect the right premiums for life insurance. More recently, it is recognized that technical testing informs scientific research just as much as hypotheses inform practical trials. The Manhattan Project itself could be seen as a case study of the mixture of engineering prowess and theoretical insights that brought about the creation of the atomic bomb.

The social contextualization of technoscience in its various communities represents a sociological rather than a philosophical designation of a changed worldview regarding the evolution of our process of acquiring, disseminating, and consuming knowledge claims. Instead of characterizing the history of ideas (including science and technology) as Kuhnian "paradigm shifts," the process is understood in terms of the social conditions (World War II and the Cold War) that led to Big Science, the military–industrial–academic complex. From this perspective, technoscience is also informed by Marxist as well as feminist critiques, since they all contribute to a fuller recognition of the inherent biases that inspire or motivate researchers in their respective communities.

Ethically, there is a significant concern that if science remains theoretical and pure, on the one hand, and technology (or engineering) remains practical and applied, on the other, the former may be exempt from ethical considerations while the latter will necessarily bear the brunt of all moral condemnation (when catastrophes erupt). But if all of these activities are combined in technoscience, all those involved will be praised or condemned depending on the consequences of their work. This designation puts added pressure on all participants to prefigure the implications of their ideas (or research), rather than presume that they work in a shielded political vacuum because of its academic prestige or industrial wealth.

As the constellation of science, technology, and engineering, technoscience is best understood in terms of the activities of a whole community, and therefore this critical approach has been supported primarily by sociologists of science. Their work should be understood as following other European movements, such as deconstruction and hermeneutics. While the latter two were concerned to find the ultimate truth hidden beneath a structural understanding of objective knowledge (Truth), the sociological turn in science studies insisted that all scientific knowledge is always already socially constructed. Any claim to objectivity or truth was false. The reductionist move to a single prism through which to look at all scientific activities was toned down in the United States with the rise of the discipline of science, technology, and society (STS) studies, which includes a stronger focus on policy implications. The relative success of

the American counterpart has been seen in government agencies that have enlisted the help of these scholars (e.g., Virginia Polytechnic Institute and Rensselaer Polytechnic Institute).

While attempting to debunk the authority of science and technology, sociologists of science may have inadvertently privileged technoscience instead. Seen as interdisciplinary in nature, technoscience is not limited to sociology but extends to other disciplines and to the critiques of science by postmodernists as well, where a plurality of perspectives and multiple interpretations are encouraged (as Raphael Sassower argues). This plurality allows us not to lose sight of the importance of science and technology in our daily life and of the different philosophical and practical aims they have. Science and technology have the potential, in concert or separately, to remain open-minded to the contribution of various quarters in the construction of knowledge, however objective and universal it may eventually become.

Raphael Sassower

See also Science and Ideology; Social Constructivism; Social Studies of Science and Technology; Sociology of Knowledge and Science; Strong Program in the Sociology of Scientific Knowledge; Technological Convergence; Transhumanism and Human Enhancement

Further Readings

Agassi, J. (1985). *Technology: Philosophical and social aspects*. Dordrecht, Netherlands: Reidel.

Aronowitz, S. (1988). *Science as power: Discourse and ideology in modern science*. Minneapolis: University of Minnesota Press.

Bachelard, G. (1934). *Le nouvel esprit scientifique*. Paris, France: Librairie Félix Alcan.

Bacon, F. (1985). *The new organon*. New York, NY: Macmillan. (Original work published 1620)

Fuller, S. (2007). *New frontiers in science and technology studies*. Cambridge, England: Polity Press.

Haraway, D (1991). *Simians, cyborgs, and women: The reinvention of nature*. New York, NY: Routledge.

Jonas, H. (1984). *The imperative of responsibility: In search of ethics for the technological age*. Chicago, IL: University of Chicago Press.

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press. (Original work published 1979)

Lyotard, J.-F. (1993). New technologies. In *Political writings* (B. Readings & K. Paul, Trans.). Minneapolis: University of Minnesota Press.

Sassower, R. (1995). *Cultural collisions: Postmodern technoscience*. New York, NY: Routledge.

Sassower, R. (1997). *Technoscientific angst: Ethics and responsibility*. Minneapolis: University of Minnesota Press.

Todhunter, I. (1865). *History of the mathematical theory of probability from the time of Pascal to that of Laplace*. London, England: Macmillan.

TELEOSEMANTICS

This entry offers an overview of a relatively recent approach, called *teleosemantics*, that has become prominent in philosophical accounts of mental representation. The entry first explains what teleosemantic theories aim at and proceeds to present varieties of such theories. Teleosemantics adds an important dimension to what constitutes representation, intentionality, and perception in humans when they relate to the outside world and can thus be seen to have fruitful applications in the social sciences, too, and in particular in cognitive psychology.

What Are Teleosemantic Theories?

Teleosemantic theories try to explain, in a way that fits into the natural sciences, the relation that meaningful mental states, such as perceptions, beliefs, and desires, have to the things they are about. They are theories of what mental representation is. Teleosemantic theories may also be extended to explain what it is for communication devices such as animals' danger signals, bee dances, and human language to represent. Distinguishing teleological theories from other naturalistic theories of representation is a certain approach to the problem of misrepresentation, a problem often overlooked by other naturalistic theories. Misrepresentation is taken to be constituted by a failure to fulfill a natural *purpose* of some kind, hence to involve teleology, though of a naturalized kind.

Consider the problem of what it is for something to be a representation but a false one—a misrepresentation. Although a false representation does seem to represent something, it does not represent anything real. How can the representation relation be a real relation yet one of its relata sometimes be unreal? If I think it is raining but it is not raining, my thought is "about" something, namely, current rain, yet there is no current rain. What then is this "about" relation that is able to obtain between a representation and something that doesn't exist?

According to teleosemantics, the answer is that representations are items that have natural purposes or *functions* that require them to represent, but these natural functions can fail to be performed. Eyes have as their function to produce sight, but they can fail in this purpose when conditions are not right, for example, when the eyes are damaged or when there is no light. Representations that are false are representations that have failed or would fail to perform their natural functions. If we were to insist on using the word *represent* in only one sense—represent—it might be said that although to fulfill their natural purposes they would have to represent, false representations do not in fact represent. They fail to represent. But the word *represent* has another use that confuses us. "It is raining," we say, means or represents the same as "Es regnet" (in German) and the same as "Il pleut" (in French), and it represents "that it is raining" (putting things in English). These are all good translations of "It is raining" (the last being homophonic). Telling what something represents is, in this sense of "represent," giving a translation. But the fact that I can give a translation of a representation into another that would represent the same thing, if it did represent, does not entail that the translated sentence *does* represent. Basic representing—representing—is bearing an actual relation of a certain kind to something in the real world. There is no need to reify unreal things for false representations to correspond to. (This way of formulating the central thrust of teleosemantics is idiosyncratic but nevertheless, perhaps, accurate.)

The natural purposes or functions invoked by teleosemanticists are taken to be derived from selection processes: natural selection, or the selection that is learning, or, in the case of language forms, social selection or meme selection. The functions that characterize representations are, intuitively speaking, survival values—what a thing or its ancestors did that ultimately accounted for its existence or continued existence. Thus, teleosemantics involves the claim that what makes an item into a representation is partly a matter of its history. Just as memories have to have the right sort of history to be real memories, so perceptions, thoughts, and sentences have to have the right sort of history to be perceptions, thoughts, and sentences. If they were to manage to serve the same sort of functions as do representations, even though they had not been designed to do so, though failing the right history, they would not be representations in the core sense. Similarly, a gadget that can nicely be used to open cans but was not designed to do so is not a can opener in the core sense. This can be supported by noting that something not designed for opening cans and hence bad at it would not be a defective can opener, nor would something not designed to represent be a false representation.

Because teleosemantic theories require that representations have the right kind of history in order to be representations, when applied to mental representations teleological theories are what are called *externalist theories of content*. They imply that the content of one's thought—what would make that thought true (or satisfied)—is not determined by anything before one's mind or within one's consciousness or even within one's head. Nor is it determined merely by current dispositions that one has to make inferences or to act.

Different Kinds of Teleosemantic Theories

Teleosemantic theories differ in what they take the functions defining representations to be, hence in what they take the representing relation to be. It has often been said, for example, both by others and by teleosemanticists themselves, that the function of a representation is to (correctly) represent or to "indicate" something else. But this tells us little, unless it is also explained what (correctly) representing or indicating is supposed to be—for example, whether representations are supposed to covary with what they represent, whether they should be lawfully caused by what they represent, or whether they must "picture" or be isomorphic to, in accordance with semantic rules of a certain kind, what they represent, and so forth. The challenge has often been construed as that of creating a description that not only characterizes all representations (at least of a particular kind) but is also not too broad, implying that certain things are representations that do not seem, intuitively, to be representations or implying that what these representations represent is not what they seem, intuitively, to represent. However, a recalcitrant problem in the background concerns just what such a theory should be expected to explain. Should it be taken as a project in conceptual analysis, an attempt to capture what the outlines of our ordinary usage of the terms *represent* and *representation* are, or perhaps of the cognitive scientist's usage of these terms? Or should it be taken as a project in theoretical science,

an attempt to describe a certain kind of important phenomenon or principle that turns up in nature in many forms and that can be appealed to for explanatory purposes? Some of the differences between teleosemantic theories can be traced to this difference in aim, the exact relevance of prior intuitions about which things are representations and of what depends, in turn, on this aim.

We can roughly classify teleosemantic theories by what they take the relation to be between a true representation and its represented. What defining function is bearing a representation relation to something else supposed to have? How does this relation come about, or how is it produced when the system that produces or employs representations functions properly? Answering this question should also answer the question of what determines what a given representation represents, that is, "what its content is."

Dennis Stampe, who may have been the first to propose a teleosemantic theory, and Jerry Fodor, who proposed one but almost immediately repudiated it as "viciously wrong," offered theories that required representations to be caused by the things they represent. For Stampe, a mental representation represented whatever caused it when it was produced by devices that were functioning properly; otherwise, its content was what probably would have produced it had it been functioning properly. Fodor claimed that various kinds of external conditions could in principle be specified under which normal perceptual/cognitive systems would operate optimally in accordance with design and that what a belief state in the brain represented was whatever would always cause it under these epistemically optimal conditions. Under these conditions, the occurrence of the represented would be sufficient for occurrence of the representation. Karen Neander proposes a causal requirement on representation, at least for ordinary perceptual representation, and for simple cases such as the impulse in the optic nerve of the frog that has often been called a "fly detector." Neander takes it to be the job of such representations to represent that which they were selected for responding to, hence that which causes them when the system operates in accordance with its design by natural selection. She recognizes that a consequence of such a causal theory is, for example, that what the "fly detector" really represents must not be flies but small, moving dark things, since these properties are the proper *causes* of the frog's response.

Fred Dretske introduced what is sometimes called an "informational" teleosemantic theory, claiming that the function of a perceptual representation is to "indicate," or to carry "natural information" about, the represented. He described a signal or representation as carrying natural information if whenever the representation is present, there is a probability of 1, in accordance with natural law, that the represented is there also. Occurrence of the representation would thus imply occurrence of the represented. A representational system was a system designed to bring this kind of relation about. Notice that this kind of relation might occur for other reasons than the representation being caused by the represented. Dretske claimed, controversially, that the representational functions of belief-like states were derived only from learning, not from natural selection.

Dretske was explicit that representations have the "function" of carrying information owing to their situation in some larger system that makes use of the information to guide behavior, thus looking beyond the causes or correlates of representations to their effects to help determine content. David Papineau and (originally) Ruth Millikan claimed that it is only the uses to which mental representations are put that are relevant to determining their content. Millikan claimed that representations emerge when two systems (or one system in two capacities) are designed to cooperate with one another, one producing representations that correspond to aspects of the world by certain (semantic) rules, the other varying its activities as guided by these representations, perhaps first in the making of inferences but ultimately always in the production of behavior, taking account of what has been represented. The emphasis here is on the relation that the interpreting or "consumer" side of the system requires between representation and the represented if it is to perform whatever functions it has in the way it was designed to.

Papineau has sometimes said that the function of a belief is to be copresent with its represented. But he also says that what a belief represents, its truth condition, is the condition that would guarantee that actions based on that belief, plus other true beliefs one has, will satisfy one's desires, the function of a desire being to produce what it represents, that is, its satisfaction condition. For both Papineau and Millikan, then, a useful "correspondence" between representation and the represented does indeed occur when the biological system functions properly,

but how this correspondence is brought about is not definitional of the representing relation.

Ruth G. Millikan

See also Cognitive Sciences; Evolutionary Psychology; Intentionality; Mind–Body Relation; Naturalized Epistemology; Social Anti-Individualism and the Mental

Further Readings

Dretske, F. (1986). Misrepresentation. In R. Bogdan (Ed.), *Belief: Form, content, and function* (pp. 17–36). New York, NY: Oxford University Press.

Fodor, J. (1990). Psychosemantics, or: Where do truth conditions come from? In W. Lycan (Ed.), *Mind and cognition: A reader* (pp. 312–337). Oxford, England: Blackwell.

Millikan, R. G. (2004). *Varieties of meaning*. Cambridge: MIT Press.

Millikan, R. G. (2005). *Language: a biological model*. Oxford, England: Oxford University Press.

Millikan, R. G. (2008). Biosemantics. In B. McLaughlin (Ed.), *The Oxford handbook in the philosophy of mind* (pp. 394–406). Oxford, England: Oxford University Press.

Neander, K. (2013). Toward an informational teleosemantics. In D. Ryder, J. Kingsbury, & K. Williford (Eds.), *Millikan and her critics* (pp. 21–40). Oxford, England: Blackwell.

Papineau, D. (1993). *The teleological theory of representation*. In *Philosophical naturalism* (chap. 3). Oxford, England: Blackwell.

Stampe, D. (1977). Toward a causal theory of linguistic representation. In P. A. French, T. E. Uehling Jr., & H. K. Wettstein (Eds.), *Midwest studies in philosophy: Studies in the philosophy of language* (Vol. 2, pp. 81–102). Minneapolis: University of Minnesota Press.

Theory of Teams

Teams are a subject of investigation in many branches of social science, and *theory of teams* has different connotations in each. In some areas, the concept of a team may be "thick," specifying features that a group must possess in order to be a team, while in others it may be "thin," referring merely to a collection of individuals. This entry will present teams in various social-scientific guises, will explain why we might consider teams to be agents, and will consider the place of teams in our social ontology.

The thinnest use of "team" occurs in economics, where it is most often used to refer to the problem of how to incentivize individual members of a workgroup who each want to maximize their own pay and minimize their effort when the employer can only observe aggregate output. The team is just a group of individuals who each has her own goals but who are all subject to the same payment schedule. Also in economics, the Marschak-Radner theory of the team is a model of individual agents with a common goal. The theory concerns how best to communicate in order to implement the optimal collective action given that agents have private information and communication is costly. Thicker still is the concept found in the management literature on organizational behavior, which explicitly differentiates teams from mere groups. A team has features such as a common goal, interdependence, and synergies between members. Researchers in this area may focus more on specialization within teams, optimal team structure, and interactions between team members. Finally, in game theory, the theory of team reasoning takes a team to be a group with a common goal and proposes a special mode of reasoning used by individuals as members of teams. This approach overlaps with the two economic theories and explains some of the phenomena discussed in the management literature.

Team Agency

In the thinnest theories of teams, the team is not an agent but merely a collection of individuals. These teams are not of much interest to philosophers of social science.

Thicker theories may allow that a team can be an agent. There are some prima facie reasons for thinking that team agency is possible. It has been proposed that teams are agents because this explains how people can solve problems of game theory and because we might attribute to a group a pattern of beliefs, desires, and judgments that are held by none of the members.

Whether or not teams are agents ultimately depends on the criteria for attributing agency. In game theory, the key characteristic of agency is *instrumental rationality*. Agents have a preference ordering that represents what they want to achieve, and they

seek to be as successful as possible according to that standard. So a minimal condition for being an agent is having an end that is pursued. In the case of a team agent, this implies that there is a group goal.

A stronger characterization of agency requires that agents have a commitment to *rational unity.* Agents are any entity that has the ability to form states that play the role of intentional attitudes, such as judgments and beliefs, and that can take steps to ensure that these states are consistent with each other. If the defining feature of agency is the ability to engage in rational deliberation, then the bounds of agency need not coincide with the individual.

These two characterizations of agency need not coincide. Under the second characterization, ephemeral groups, such as two strangers pushing a car, will not exhibit agency. But these groups have a common goal and hence may be agents according to the theory of team reasoning.

Ontological Status of Teams

A related but separate question is the ontological status of teams. In the early 20th century, an *emergentist* tradition, inspired by Georg W. F. Hegel, considered teams to be transcendental entities over and above their individual members. Advocates of this approach were happy to countenance the possibility of "group minds" and tended to support more holistic explanations in social science. Others have found the idea of a group mind mysterious or problematic.

More recently, the dominant paradigm has been *eliminativist,* asserting that groups are reducible to individuals and that any property ascribed to a group can be reexpressed in terms that refer solely to the individuals who constitute it. Eliminativism is associated with methodological individualism in philosophy of social science, which denies that groups are a part of good explanations of social phenomena.

A middle road has been offered by Christian List and Philip Pettit, who argue that groups "supervene" on their members. The notion of supervenience is familiar from other areas of philosophy. As applied to teams, the idea is that facts about the individual members determine facts about the team. There is no additional factor required for the constitution of a group agent, over and above those about the collection of individuals who constitute the team. However, the relation between the agency of the individual members and the agency of the group may be very complex, and features of the group agent are not easily reducible to those of its members. Hence, talk of team agency is nonredundant, and teams may be a part of good social-scientific explanations.

Natalie Gold

See also Collective Agents; Collective Goals; Collective Rationality; Collective Values; Emergence and Social Collectivism; Group Beliefs; Group Mind; Holism, in the Social Sciences; Individualism, Methodological; Judgment Aggregation and the Discursive Dilemma; Supervenience; Team Reasoning

Further Readings

Bacharach, M. (2006). *Beyond individual choice: Teams and frames in game theory* (N. Gold & R. Sugden, Eds.). Princeton, NJ: Princeton University Press.

Gold, N. (Ed.). (2004). *Teamwork: Multi-disciplinary perspectives.* London, England: Palgrave-Macmillan.

Gold, N., & Sugden, R. (2008). Theories of team agency. In F. Peter & H. B. Schmid (Eds.), *Rationality and commitment* (pp. 280–312). Oxford, England: Oxford University Press.

List, C., & Pettit, P. (2011). *Group agency: The possibility, design and status of corporate agents.* Oxford, England: Oxford University Press.

Marschak, J., & Radner, R. (1972). *Economic theory of teams.* New Haven, CT: Yale University Press.

Theory Theory

This entry introduces one of the alternative theories of folk psychology, the so-called Theory Theory, presents its main features, and reviews criticisms of it.

The main competitor of the Theory Theory is the *Simulation Theory.* They are both alternative accounts of what are known as ToM, that is, *theories of mind:* theoretical accounts explaining how people ordinarily understand each other's beliefs and actions in everyday settings by ascribing psychological states to them, without any prior scientific schooling in the notions people employ in doing so. The Theory Theory is thus a philosophical-psychological theory about a "theory" people apply in order to understand each other—a kind of mind reading.

The Theory Theory holds that knowledge of a (certain) "theory" helps us (the "folk") understand, predict, and explain the actions of others in terms of their psychological properties, such as the beliefs, desires, and intentions we ascribe to them. For instance, we explain why George went to the tavern by reference to his wanting a beer and believing that he could get one in the tavern. The theory has been very influential both in philosophy and in psychology.

In developmental psychology, a Theory Theory approach has been championed by Alison Gopnik and Andrew Meltzoff in relation to the child's cognitive development. The child is seen as a "scientist in the crib" constructing a "theory" about the world. The basic idea is that children do so by using the same cognitive tools that adult scientists use in constructing scientific theories. Children develop theories enabling them to predict outcomes and others' actions and assess evidence, and they even "experiment" and "test" in exploring the world around them. Thus, we grow up developing and employing a theory to understand others.

What Does the Theory Theory Hold?

The motivation behind the Theory Theory is the following problem. Though we are familiar with what people do and say and how they express themselves, we never have any *direct access* to the internal psychological states that we assume cause their observable behavior. It therefore stands to reason that we *infer* the existence of such mental states. But to do that, we must know which psychological states go with what behavioral manifestations, hence the need for a *theory*.

Many have thought that the situation here is no different from what we are faced with when attempting to understand natural phenomena. We need something like a scientific theory to understand others. In physics, we often posit so-called unobservables to explain the behavior of observable objects. For instance, we can observe the effects of gravity on how objects move but not gravity itself. It is sometimes thought that the same is true of mental states. Because we cannot observe them in others, but nevertheless assume that they cause their behavior, they gain a status similar to that of unobservable entities in scientific theories.

Problems With Theory Theory

One problem with thinking about mental states as *unobservables* is that most people believe that they have direct access to their (own) mental states. If they do, it is hard to see how they could be unobservables. Many theory theorists, however, think we do *not* have unmediated access to the contents of our own minds. Only by learning about other minds do we get to understand our own, suggesting that our knowledge of our own minds is *mediated*. Some of the developmental literature in psychology supports this idea. Not all theory theorists deny, however, that we have more direct access to our own mental states than those of others.

When queried, most people (including philosophers) are unable to produce anything like the *law-like* generalizations that the Theory Theory leads us to expect that they know. Most theory theorists, in fact, maintain that our knowledge is *tacit,* much like our knowledge of grammar in a Chomskian framework. The assumption is that we can know a theory and apply it without having conscious access to it and with it playing only a limited inferential role in our general reasoning. There are issues with this interpretation, however. We consciously know much more about the principles of psychology than we know of Chomsky-style grammar, *and* we seem to possess, and use freely, the concepts that are involved in folk psychology but not grammar. We have, and use, concepts such as belief, intention, desire, and fear freely but not artificial concepts such as *wh*-traces, a technical item of the complicated Chomskian theory of generative grammar.

The classic version of the Theory Theory holds a Hempelian-Nagelian view of scientific theories, according to which a scientific theory is a coherent body of knowledge containing theoretical terms that refer to unobservable entities or properties that are defined in terms of one another and are related in terms of laws. The theory provides explanations and predictions in terms of these laws. Using such a picture of our folk psychological knowledge, Paul M. Churchland has argued that our standard ToM is false and should be discarded like any other false theory. His reasons are that the theory is stagnant, insofar as it is no different from that held by the Ancient Greeks; it is explanatorily deficient—it fails to explain important psychological phenomena such as mental illness, creative imagination, and learning—and it shows poor prospects of being integrated into the natural sciences, none of which operate with intentional properties, such as believing that *p* or desiring that *q*. People have responded that folk psychological theory is the only game in town and

that although parts of it may be false, it is unlikely to be false in its essentials.

Alternatives

Other versions of the Theory Theory suggest that rather than think of our knowledge of folk psychological theory in terms of knowledge of lawlike relations, we should think of it as knowledge of *models*. Models are like exemplars, which we apply to real-life situations, assuming that the principles that guide the workings of the model apply to reality in certain respects and to a certain degree. We often use ideas of what ideally rational agents would do to understand other less rational agents.

Daniel Dennett's so-called intentional stance may be understood as a version of the Theory Theory. Taking the intentional stance on others is to *assume* that they have the beliefs and desires that it is rational to have—in other words, that they have intentional states. Doing so helps us do explanatory and predictive work that purely physical or functional frameworks cannot provide. Dennett believes that mental states as we conceive of them do not exist, but he thinks we need to think of others in intentional terms nevertheless. By contrast to both Churchland and Dennett, Jerry Fodor has argued that we should expect any future science of mind to retain the central categories of folk psychological theory. So far, both psychologists and philosophers have continued to work with categories that are essentially those of folk psychology.

Heidi L. Maibom

See also Folk Psychology; Laws of Nature; Mind–Body Relation; Models in Science; Simulation Theory; Tacit Knowledge

Further Readings

Carruthers, P., & Smith, P. K. (Eds.). (1996). *Theories of theories of mind*. Cambridge, England: Cambridge University Press.

Churchland, P. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78,* 67–90.

Dennett, D. C. (1987). *The intentional stance.* Cambridge: MIT Press.

Fodor, J. A. (1987). *Psychosemantics*. Cambridge: MIT Press.

Gopnik, A. (2011). *The philosophical baby: What children's minds tell us about truth, love and the meaning of life.* New York, NY: Farrar, Straus, & Giroux.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories.* Cambridge: MIT Press.

Gopnik, A., Meltzoff, A. N., & Kuhl, P. (2001). *The scientist in the crib: What early learning tells us about the mind.* New York, NY: Perennial.

Maibom, H. (2003). The mindreader and the scientist. *Mind & Language, 18,* 296–315.

Therapy, Psychological and Philosophical Issues

The entry discusses therapy as understood by psychology, in the form of psychotherapy, and concentrates on the two main kinds of philosophical issues surrounding it: first, the issue of values involved in psychotherapy, or in general the normative challenges unavoidably related to it, and, second, the issue of its epistemic status, that is, the question of the evidential support the outcome of psychotherapy may be said to enjoy.

At one time, only the so-called insight therapies, those heavily influenced by the views of Sigmund Freud, were classified as psychotherapies. Today, the term *psychotherapy* is used in a wider sense to include any sort of psychological therapy, including the behavior therapies and the cognitive therapies. In this wider usage, there are more than 400 different types of psychotherapy. The question of which forms of psychotherapy are effective, for which clinical disorders and under what conditions, has been an enduring topic of controversy.

At first appearance, the issue of effectiveness is straightforwardly empirical. We are asking about a causal connection between treatment and outcome. How else are we to answer this causal question except through empirical inquiry? Yet, when we press further and ask about the kind of empirical evidence that is needed and what the evidence shows so far, various deeper questions emerge, some of which are clearly philosophical.

The Values Question

The picture of psychotherapy accepted by many is of an enterprise that is through and through *value-laden*. It is not merely that psychotherapists sometimes make ethical judgments in their practice but that such judgments are everywhere and are

unavoidable. Here is one illustration of this view discussed in the literature.

Bob is suffering from anxiety and depression. Part of his problem is that his wife has moved in with a much younger man. Bob wants help, and he wants advice. The wife's new relationship is not working out, and there is a good chance that she is willing to go back to her husband. Bob, apparently seeking moral advice, asks the therapist if he should forgive his wife and if he should try to work out the problems in their marriage.

As the therapy progresses, the therapist must make more ethical decisions in deciding the goals of the therapy and the best means to achieve them. The therapist decides that the therapeutic goal should be for Bob to gain insight into the unconscious and childhood roots of his current difficulties, but Bob just wants to be less confused, less anxious, and less depressed. The insurance company, on the other hand, wants the focus to be on a quick, cost-effective reduction in symptoms. Further ethical decisions must be made when the therapist realizes that others, including Bob's children, have a stake in the therapeutic outcome. The therapist reasons that he must make an ethical decision concerning the weight of their interests compared with those of the client and the insurance company.

Some who see the Bob case as a model for psychotherapy reach provocative conclusions such as this: Psychotherapy research and practice are essentially expressions of moral principles, the practice of psychotherapy is not in any way separable from the practice of ethics, and the psychotherapist is someone working in applied ethics.

Many cite the value problem as the main impediment to developing a natural science account of psychotherapy, but they need another premise. It is not enough to say that psychotherapists must make value judgments in their research and practice. That is consistent with a natural science view of the field. The needed additional premise is that there is no way to establish the truth or falsity of value judgments.

The issue of explaining how value judgments can be proved is a difficult one (and is, of course, central to philosophy in general), but what might help is to split the value problem in two, reflecting the distinction philosophers draw between theories of the nonmoral good and theories of the morally right. Traditional hedonism, for example, treats pleasure as the only intrinsic good, but it is silent on the moral issue of whether and under what conditions agents ought to seek their own pleasure. In contrast, Kantianism and Mill's utilitarianism do talk about what is morally right or wrong.

The evaluative questions that arise in psychotherapy's outcome research are generally not moral questions; they are questions about whether treatment effects are beneficial, neutral, or harmful to individuals. Moral issues, when they do arise, usually come up in choosing therapeutic goals or when clients seek moral advice from a therapist.

Although not everyone agrees to this, explaining how a treatment effect can be shown to be beneficial for an individual may be easier than showing how moral judgments can be proven true or false. Assuming that Bob is not being irrational or misinformed in wanting to be less depressed, it seems, at least before we wade into the philosophical issues, that one can make the case that satisfying his desire to be free from depression would be a benefit for him. Some philosophical theories try to explain how judgments about what is beneficial to someone can generally be warranted by linking beneficial outcomes to the satisfaction of an agent's desires, provided that certain conditions are met, such as the absence of irrationality and relevant ignorance.

There are likewise cognitivist moral theories that try to show how moral judgments can be proven or disproven. If one of these can be warranted, then the moral component of the values problem can also be handled; but if moral skepticism is true and no moral judgment can be proven, then we might try a different option and argue that the extent to which psychotherapists must make moral judgments, at least of a categorical kind, is greatly exaggerated.

Disagreements about the proper goals of therapy are often about likely consequences and not about moral issues. For example, the debates between psychoanalysts and behavior therapists about making symptom remission the goal of therapy has mainly been about the symptom substitution hypothesis, with psychoanalysts contending and behavior therapists denying that the elimination of symptoms without a resolution of underlying unconscious conflicts generally causes worse symptoms to develop. The Bob case is fictional, but if his therapist is appealing to the symptom substitution hypothesis or merely claiming that the most efficient way of solving Bob's problems is to gain insight into their underlying cause, then he is making an empirical, not a moral

point. If, rather oddly, the therapist were making the moral claim that it is better on moral grounds to seek insight into unconscious causes, then the therapist's behavior would be atypical and could not be reasonably taken as a model of what psychotherapists generally do.

Even where therapists give moral advice, and even if we count this as part of the psychotherapy and not something merely added on, the judgment is often conditional, not categorical. If the therapist were to tell Bob that he ought to reconcile with his wife, the therapist might well mean that Bob should do this on condition that he wants to be happy or wants his children to be happy. Conditional moral judgments of this sort can often be proved or disproved empirically. The question of whether Bob's reconciling with his wife will increase his happiness is a causal question that can be answered by finding the right sort of empirical evidence.

Psychotherapists do sometimes give categorical moral advice—"You ought to do this no matter what your goals are"—but it remains to be shown that this fact creates a serious problem for the general practice of psychotherapy or for a natural science approach to researching its effects.

Standards of Evidence

Many of the controversies about what has, and has not, been shown to be effective reflect different approaches to standards of evidence. The issues are partly empirical; but when disagreement extends to basic standards, they are also partly philosophical.

Many psychotherapists have followed Sigmund Freud in relying on uncontrolled clinical case studies. The idea is that by the time the therapy is terminated, the therapist can often see that the treatment has brought about a reduction in symptoms or some other beneficial result. The standard objection to this idea is that at most the therapist can determine by observing the patient that improvement has occurred, but she cannot know the cause of the improvement without ruling out a credible rival hypothesis: Spontaneous remission has taken place, which means that events occurring outside the therapist's office were causally responsible for the beneficial outcome. Even if this can be ruled out, there is the problem of discounting another credible rival hypothesis, that the main curative element was not the therapy but the patient's belief that the treatment would be effective—in short, that there was a placebo effect.

These twin problems also arise when pharmaceutical treatments are studied for their effects on psychological problems such as depression or anxiety. The standard way to deal with them is to use randomized clinical trials with a credible placebo. However, many psychotherapists object to such trials on ethical grounds—patients in the placebo group are deceived about the treatment they are receiving—or because they believe that they are unnecessary.

Whatever the merits of these objections, it still has to be explained how credible rivals to a hypothesis of therapeutic effectiveness can be ruled out without employing randomized clinical trials or how we can obtain credible evidence without ruling out credible rivals. The first option has been explored recently in the medical literature, where some have argued that certain types of observational studies, such as sophisticated cohort and case–control studies, can take the place of randomized clinical trials, but neither of these sorts of trials has been widely used in psychotherapy research.

If, contrary to what many psychotherapists believe, randomized clinical trials are ultimately necessary for establishing effectiveness, then as of now, most of the 400 or so psychotherapies have not yet been shown to be effective. As more and more randomized clinical trials are carried out, and providing they are of high quality, the quality of the outcome evidence will improve.

But there is another problem: the *integration problem*.

The best evidence that a medical treatment is generally effective for a certain malady is a megarandomized clinical trial with 50,000 or more subjects. No such trial has ever been done in the field of psychotherapy. When smaller trials are done, it is very common, partly because of the great variability in responses to treatments, for the results to disagree. Some studies show positive results, some negative results, and some neutral results. This gives rise to the integration problem: the problem of putting all of the evidence together, analyzing it, and reaching a reasonable causal verdict.

The standard way of dealing with this problem in psychotherapy and medicine is to use *meta-analysis*. A statistician collates the results of all the relevant studies and calculates an "effect size," assuming

that there is a quantitative measure of a therapeutic outcome and at least one control group, by subtracting from the mean score for the treatment group the mean score for the control group and dividing the result by the standard deviation for the control group. An average of the effect sizes for all the studies is then calculated and is said to be the average effect size for the treatment. This may work if all the studies are high-quality randomized clinical trials, or some adequate substitute, but in much of the psychotherapy outcome research, the studies are uniformly inferior, or there is a mixture of good and bad studies. In such cases, doing a standard meta-analysis will not by itself permit the inference that a therapy is effective.

One possible solution tried in medical research, but not in psychotherapy research, is to develop a quantitative system for rating design excellence. Many such systems have now been developed, but the philosophical problem of validating the numerical values for a system has so far not been solved. The upshot of this is that even where psychotherapy has been the subject of experimental studies, there is often good reason not to trust the meta-analyses that have been done to integrate the conflicting evidence and reach a causal verdict. The main exception is where all the trials are of high quality and the results are in agreement. This, unfortunately, is not the standard case in psychotherapy research.

Edward Erwin

See also Collective Identity and Cultural Trauma; Experiments in the Social Sciences: Ethical Issues; Evidence-Based Policy; Moral Cognitivism; Personal Identity and Trauma; Psychoanalysis, Philosophical Issues in; Schizophrenia: Psychoanalytic, Phenomenological, and Contemporary Philosophical Approaches

Further Readings

Erwin, E. (2006). Randomized clinical trials in psychotherapy outcome research. *Philosophy of Science, 73,* 135–152.

Erwin, E. (2010). *Behavior therapy: Scientific, philosophical, and moral foundations.* New York, NY: Cambridge University Press. (Original work published 1978)

Erwin, E. (2011). Evidence-based psychotherapy: Values and the a priori. In M. Veber & M. Shaffer (Eds.), *What place for the a priori?* (pp. 33–59). Chicago, IL: Open Court.

Goodheart, C., Kazdin, A., & Sternberg, R. (2006). *Evidence-based psychotherapy: Where practice and research meet.* Washington DC: American Psychological Association.

Miller, R. (2004). *Facing human suffering: Psychology and psychotherapy as moral engagement.* Washington, DC: American Psychological Association.

Smith, M., Glass, G., & Miller, T. (1980). *The benefits of psychotherapy.* Baltimore, MD: Johns Hopkins University Press.

Tjeltveit, A. (1999). *Ethics and values in psychotherapy.* New York, NY: Routledge.

Thought Experiments

There is no agreed definition of the term *thought experiment*. However, two important strands of usage may be discerned. In one usage, *thought experiment* refers to imaginary or hypothetical examples used to bolster, illustrate, or convey a theory. In another, narrower usage, it refers to the use of such examples to *test* a theory and thus provide evidence for or against it. There is no need to choose between these usages, but it is fruitful to keep them apart. This entry will discuss each notion in separate sections. For brevity, the term will be understood below in the sense indicated by the respective headings.

Thought Experiments as Examples

Sometimes, imaginary examples are used simply to illustrate an abstract theory or concept in a textbook or a classroom setting in economics, jurisprudence, or the sciences. Their utility for this purpose is due to the fact that they are free to abstract from messy, distracting detail and to idealize away nonsalient causal factors that would always be interfering in an actual case.

Thought experiments used for this purpose may draw on analogies or use extrapolation. Albert Einstein and Leopold Infeld illustrated the equivalence of gravitational and inertial mass with the famous case known as *Einstein's elevator,* in which a person will observe light bending whether the

elevator is accelerating in an inertial frame or subject to gravitation. Isaac Newton imagined a cannonball's trajectory gradually widening or extending to the point where it doesn't come down at all, thus illustrating how gravitation keeps the moon in orbit.

Thought experiments may also be used to bolster a theory—without, strictly speaking, testing it—by showing its explanatory potential. Darwin had no access to the historical evidence that would go into a detailed explanation of the evolution of biological variation. But in *The Origin of Species,* he invites the reader to imagine various scenarios in which selection would lead to what we now observe, thereby establishing his theory as a *potential* explanation, contrary to creationist objections. Such use of thought experiments to show a theory's explanatory merits is not sharply distinguishable from mere illustration. In both uses, the point of the exercise is to see what would happen in the scenario on the assumption that the deployed theory is true.

Thought Experiments as Tests

Sometimes, however, hypothetical cases are invoked to check whether a theory yields the right result in them. This outcome, in turn, is taken to indicate whether the theory is true. This use of hypothetical cases for providing evidence is very common in philosophy, but it also occurs in the sciences. Galileo asked what would happen if a cannonball and a musket ball were connected by a string and then dropped. In Aristotle's theory, the speed of a falling body is proportional to its weight. Galileo argued that, in his hypothetical scenario, Aristotle's theory implies that the composite body should fall both faster than the cannonball on its own (since it is heavier) and slower (since the small musket ball should act as a drag), which is absurd.

In ordinary testing by experiment—in a wide sense, covering not only laboratory experiments but also observational testing in, for example, astronomy and sociology—a theory is tested by its logical consequences: If the theory is true, certain observations will be made given certain initial conditions. The experiment then proceeds by establishing that the initial conditions hold and checks whether the theory's predictions are right. (This simplifies matters a bit.)

Correspondingly, thought-experimental testing proceeds by establishing that a hypothetical scenario is possible and checks whether the theory is correct about what would be the case in that scenario. In conventional experiments and thought experiments alike, an outcome at odds with the theory counts as a prima facie counterexample, indicating that the theory is false; an outcome in accordance with the theory may be some evidence that it is true. Thought experiments parallel ordinary experiments but differ in being *modal*—that is, in dealing with possible, rather than actual, cases.

The modal character of thought experiments arguably brings the parallel with ordinary experiments to a halt. For thought experiments, three questions loom: (1) How can merely possible cases be evidence? (2) How do we know whether a scenario is possible, and what would be the case in it, given that we do not empirically observe this? (3) What is the relation between the particular test case and the theory that is tested?

These questions are interrelated. Consequently, various attempts to answer any of them tend to either address or presuppose certain answers to all three. Some theorists take the epistemological puzzle raised by Questions 1 and 2 to motivate skepticism about thought experiments (they do not give evidence) or reductionism (thought experiments give knowledge but may be reconstrued as general arguments that do not invoke case particulars). Others take it to motivate Platonism or rationalism, according to which thought experiments employ intuitions, understood as a nonperceptual, a priori analog of observation. Some argue that although modal and particular, thought experiments don't require any peculiar a priori cognitive capacities. Still others hold that the use of thought experiments as tests should be restricted to the testing of distinctly a priori theorizing.

Advocates of these various positions tend to adduce very different instances as examples bolstering their own interpretation. In discussions of thought experiments in the philosophy of science, Galileo's case has become a standard instance of a successful thought experiment. Recent debates of thought experiments in philosophical methodology often treat Gettier cases (undermining the standard definition of knowledge as justified true belief) as paradigmatic. The latter debate is currently engaged in the question of what the content of thought-experimental claims (or intuitions) is.

Various suggestions seem to accommodate the wish to render thought experiments epistemically respectable by construing the content as something both plausibly accessible to intuition and relevant to the theory under testing. This has in turn led many theorists to construe the tested theory as a (universal) necessity claim. But it is worth noting that when historians invoke hypothetical scenarios—alternative histories—to argue for or against claims of causal relations, they seem to be engaged in thought experimentation as testing. Yet the theories they test by means of thought experiments, while modal, seem to be making neither universal nor necessity claims.

Recently, experimental philosophers have challenged the very appeal to intuitions as evidence. Current work—both experimental and theoretical—on intuition should prove relevant to the status of thought experiments as tests.

Sören Häggqvist

See also A Priori and A Posteriori; Epistemology; Experiment, Philosophy of; Experimental Philosophy; Experiments in Social Science

Further Readings

Brown, J. R. (2010). *Laboratory of the mind: Thought experiments in the natural sciences* (2nd ed.). London, England: Routledge.

Häggqvist, S. (2009). A model for thought experiments. *Canadian Journal of Philosophy, 39,* 56–76.

Horowitz, T., & Massey, G. (Eds.). (1991). *Thought experiments in science and philosophy.* Lanham, MD: Rowman & Littlefield.

Nersessian, N. (2007). Thought experiments as mental modelling: Empiricism without logic. *Croatian Journal of Philosophy, VII,* 125–161.

Sorensen, R. (1992). *Thought experiments.* Oxford, England: Oxford University Press.

Williamson, T. (2007). *The philosophy of philosophy.* Oxford, England: Oxford University Press.

Time, Philosophical Theories of

The nature of time has puzzled philosophers throughout history. Time pervades our lives and experiences. The notion of an indeterminate or "open" future contrasting with a determinate or "fixed" past has seemed essential to certain views of free will, and more generally, our views about time are likely to be central to our views about reality as a whole. Yet there is something deeply puzzling about time. The puzzle is usually thought to have something to do with the notion that time *passes*. This entry discusses philosophical theories concerning the passage of time, along with some related views about the existence of the past and future.

McTaggart and the A and B Theories of Time

Current theories of time are divided into two varieties, known as *A theories* and *B theories*. These names are derived from J. M. E. McTaggart's notions of the *A series* and *B series* of time. Let us use the term *A property* to denote temporal properties such as *being present, being past to some degree,* or *being future to some degree*. Then an A series is a time series in which times are ordered according to their A properties. One time is present, and all other times are past or future to differing degrees. As time passes, a different time becomes present, and a new A series arises; or to put it another way, the position of each time along the A series constantly *changes*. In a B series, by contrast, times are arranged in an order, but no time is past, present, or future, and nothing about the B series ever changes. Instead, times are ordered according to the relations in which they stand, eternally, to one another—relations of being *earlier than, later than,* or *simultaneous with*.

McTaggart held that only the A series truly described time; change, he held, was essential to time, and only where there are A properties can there be genuine change. Yet McTaggart held the A series to be paradoxical. Because time passes, every time has its turn at being present and, similarly, every time has its turn at being past and being future. Yet no time can be past, present, and future; these properties are incompatible and hence the paradox. The obvious response is to accept that every time has every A property but to deny that any time is past, present, and future all at once. Instead, the year 2100, for example, is *presently* in the future, it *will be* present, and it *will be* past. So a given time has only a series of compound "tenses," such as *presently future, future past,* or *past past future;* it does not have the simple properties *past, present,* and *future* all together.

Much debate has arisen over whether the move to compound tenses removes the paradox; McTaggart

thought not. McTaggart's own radical conclusion was that since, according to him, there can be no real A series, time itself is unreal. During the latter half of the 20th century, however, a number of philosophers have drawn a quite different conclusion from McTaggart's paradox. Advocates of the B Theory agree with McTaggart that there is no real A series, but they hold that real time is adequately described by the B series. Thus, they hold that no time is really past, present, or future, and along with this, they deny that there is any real passage of time. All locations in time have the same metaphysical standing, much like locations in space. Just as no single location in space is really, objectively *here* (it is only said to be "here" by a person at that location), similarly no location in time is really, objectively *now*. Some B theorists are motivated by McTaggart's paradox, but others have arrived at the view for independent reasons. Advocates of the A Theory, by contrast, hold that there is a real A series and that time passes.

Versions of the A Theory

A theories and B theories come in a number of different versions. Most B theories are broadly similar, differing mainly in what they say about the relation between time and space. The B Theory holds that all times are equally real; the only reason why the present seems special to us is that we happen to be located there. Others (or other temporal parts of us) are located at other times, which they describe as "now" with equal justification.

Different A theories, however, are associated with very different views about reality. The traditional "moving-spotlight" A Theory holds that all times are equally real but the present "moves" along the time series much like a spotlight moving along a line, illuminating one location after another. The "growing-block" theory, by contrast, agrees that the past and present exist but denies that the future exists. According to this theory, reality consists of a space–time "block" consisting of all past and present entities but no future entities, and the block grows as time passes and more of reality comes into being. Hence, the past is "fixed," whereas the future, which is yet to come into existence, may be seen as "open" (insofar as the laws of physics allow for different possible futures). Another theory, sometimes known as the "shrinking-tree" theory, agrees that a single past and present exist but holds that all possible futures also exist. As time passes, a single state of the world comes to be present, then passes into the past. Reality is thus like a tree whose lowest branches continually fall off as time passes.

Currently, however, the most popular version of the A Theory is *presentism,* according to which reality consists only of the present. The past and future do not exist. But the nature of the momentary reality constantly changes as time passes. Strictly speaking, presentists need not accept that there are A properties at all; there are no past or future times that could instantiate properties of pastness and futurity, and presentness is therefore redundant. Many presentists do hold nonetheless that there are irreducibly tensed facts about the present (e.g., the present fact that there were dinosaurs in the past); and almost all presentists hold that time passes.

One advantage of presentism is that it clearly avoids McTaggart's paradox. If reality consists of a single present time, then no time ever instantiates pastness or futurity, so no paradox arises. But presentism also faces problems. The problem that has received the most attention is that of explaining what makes past or future tensed utterances true. Consider, for example, the claim that there were once dinosaurs. If the past exists, then the claim that there were once dinosaurs is made true by the dinosaurs themselves, located in the past; as philosophers sometimes say, the dinosaurs are the *truthmakers* for the claim that there were once dinosaurs. But if there is no real past, then there are no truthmakers for claims about the past; so presentists must either deny that all truths require truthmakers or else find an unexpected truthmaker residing in the present.

Fatalism

If, as presentists and growing-block theorists hold, the future does not exist, this makes it easier to defend the view of free will, according to which the future is open and our free choices help determine which of the possible futures comes to be. But if, as B theorists and moving-spotlight theorists hold, there is already a single determinate future, just as fixed as the past, then, on the face of it, this notion of free will must be abandoned, for there is only one possible future course of events (this is the view sometimes known as *fatalism*). Views about the metaphysics of time thus tend to constrain views about free will,

and vice versa. It is important to note, however, that the problem of reconciling free will with an already existing future is distinct from the problem of reconciling free will with determinism, the view that the current state of the world, combined with the laws of nature, determines all future events. Determinism could be true even if presentism were true; for example, the fact that no future yet exists does not entail that there are many possible ways for the future to turn out.

Problems for the Passage of Time

The notion that time *passes* is at the root of the disagreement between the A theorist and the B theorist. The notion of passage is at once familiar and elusive; we think we know from experience what it means for time to pass, yet it is notoriously hard to explain what this amounts to. We seem only to be able to describe it using puzzling metaphors such as *passage* or *flow*.

These metaphorical descriptions have led to challenges. One of these concerns the *rate* at which time passes. It is sometimes claimed that if time passes then it must pass at some rate. But at what rate does time pass? A standard answer is that it passes at the rate of 1 second/second. There has been much recent debate over whether this is a possible rate. Moreover, even if we can make sense of a rate measured in seconds per second, it sounds worryingly as though there would have to be two distinct time series for the rate of time's passage to be a ratio between two quantities measured in seconds. But the notion that time passes is clearly not equivalent to the claim that there are two different time series, the units of which stand in a one-to-one ratio to each other. Much clarification is needed here.

It is open to the A theorist to reject these puzzles as simply taking the metaphors too literally. Perhaps, they may say, the notion of passage is indeed rather mysterious and cannot easily be explained in terms of anything else, but this is not an adequate reason to reject it; for these are mere quibbles when set against our immediate familiarity with the passage of time in experience. The B theorist must hold that our apparent experience of time passing is an illusion and must therefore explain the nature of the illusion and the reasons why it occurs. The A theorist, it is assumed, has no such problem with experience. This has often led to an assumption that the A Theory is the default theory and the B Theory is a counterintuitive theory for which strong arguments must be given.

Recently, however, some philosophers have challenged the assumption that experience gives us any reason to accept the A Theory. If we experience time passing, then presumably the passing of time must have a role in making our experience the way it is. Yet when one starts to consider possible mechanisms whereby the passage of time could bring about the experience of it, problems loom. All other kinds of experience seem to involve a unique causal chain leading from the experienced phenomenon to the experience itself; at any rate, there must be something that makes it the case that the experience and the phenomenon are uniquely related to one another. The experience is of passage, not of another phenomenon; and no other kind of experience is an experience of passage. Giving an account of what makes this the case promises to present a significant challenge to the A theorist.

Simon Prosser

See also Determinism; Free Will, Philosophical Conceptions of; Libertarianism, Metaphysical; Metaphysics and Science; Personal Identity and Trauma; Time, Social Theories of

Further Readings

Le Poidevin, R. (Ed.). (1998). *Questions of time and tense.* Oxford, England: Clarendon Press.

Le Poidevin, R., & McBeath, M. (Eds.). (1993). *The philosophy of time.* Oxford, England: Oxford University Press.

McTaggart, J. M. E. (1908). The unreality of time. *Mind, 18,* 457–484.

Mellor, D. H. (1998). *Real time II.* London, England: Routledge.

Oaklander, L. N., & Smith, Q. (Eds.). (1994). *The new theory of time.* New Haven, CT: Yale University Press.

Price, H. (1996). *Time's arrow and Archimedes' point.* Oxford, England: Oxford University Press.

Prosser, S. J. (2011). Passage and perception. *Noûs.* doi:10.1111/j.1468–0068.2010.00816.x

Smart, J. J. C. (1949). The river of time. *Mind, 58,* 483–494.

Williams, D. C. (1951). The myth of passage. *Journal of Philosophy, 48,* 457–472.

Time, Social Theories of

Time is our self-conscious awareness of change. We can measure such changes by means of the shadow that moves across a sundial, the sand that falls through an hourglass, the hands that traverse the face of a clock, or the vibrations of a cesium atom. Human societies create divergent ways to understand time, and in turn, the rhythms of activity in these societies are shaped by their respective systems of time reckoning. Social theories of time attempt to explain or account for some aspect of this reciprocal relationship between time and society. For all human beings, however, death is a fact of life; we know our days are numbered. The essence of human existence is temporal because our consciousness of impending mortality brings about a uniquely human concern with time itself. Consequently, theories that address time or temporal experience are important issues in the humanities and social sciences. These theories form two principal lines of inquiry: (1) *the social construction of time* and (2) *variation in temporal experience.*

The Social Construction of Time

The predominant theoretical perspective emerged at the beginning of the 20th century with the idea that time is a *social institution.* Like religion, family, and government, time is a social institution because it represents a culturally specific solution to one of the challenges that confront all human societies: the temporal organization of social rhythms in our behavior. Systems of temporal organization vary historically and culturally. Unlike hunting and pastoral peoples, for example, agrarian peoples typically establish a regular day of rest. Time reckoning is also variable. The people in one society may agree to meet at a certain hour, but in a society without clocks, people may agree to meet when the sun is at a certain point in the sky. The people in one society may estimate intervals of time in terms of minutes or seconds, while those in another society use the time it takes to cook rice or fry crickets.

The social construction of time entails the *creation* of temporal systems that are products of human ingenuity and artifacts of social interaction. These temporal systems have histories, and they are *culturally relative;* they are neither natural nor inevitable. Yet they are also real or objective features of cultural arrangements, and once they have been established, it is very difficult to alter them. Indeed, subsequent to their establishment, these temporal systems act back on the individuals who enact them with exteriority and constraint.

The 7-day week is a prime example of these social conventions. It is so familiar that we take it for granted, but its arbitrary origins are rooted in the seven "planets" of Babylonian astrology and the creation stories of the ancient Hebrews. Nonetheless, the week structures the rhythm of our activity. Each day of the week seems to have intrinsic qualities, yet these characteristics quickly evaporate when we go on vacation or extraordinary circumstances free us from our usual schedule. The 7-day week is certainly not universal. Cross-cultural research reveals that various societies have had weekly rhythms that are 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 19, or 20 days in length. Those who espouse radical philosophies have tried to abolish the 7-day week because of its religious roots. Following the French Revolution, the new regime attempted to enforce a 10-day week, and 140 years later, Soviet Bolsheviks tried to establish a 5-day week. Each of these efforts lasted for more than a decade, and each of them offered a feasible alternative, but the people of both societies rejected these revolutionary temporal systems because of their cultural commitment to the traditional 7-day rhythm.

The international standardization of time is another socially constructed temporal system. The decision to place the prime meridian at Greenwich, England; the number and width of the time zones; the location of the international dateline in the Pacific Ocean—these are social conventions negotiated at an international conference during the latter part of the 19th century. Although it is not easy to revise temporal conventions, they are subject to change. In the 20th century, for example, the international community agreed to redefine a second. The older (but now too imprecise) definition was one 86,400th of 1 day. As of 1967, however, 1 second is defined as exactly 9,192,631,770 oscillations of the cesium atom.

The social distribution of time is a function of relative power and status. If we define time as a scarce and valuable resource, then we can see how waiting and delay are structured by a temporal system

of inequality. We are willing to wait for what we want, and the more we want it, the longer we are willing to wait. It follows that powerful and prestigious people demand that others wait for them as a form of deference and avoid waiting by hiring others to stand in lines for desired goods and services. An emerging *theory of temporal politics* elaborates on these principles to explain how ruling regimes in Latin America construct their own authority by making citizens wait interminably for various governmental services. A related theory concerns postindustrial nations, where politics and economics are increasingly desynchronized because the legislative processes of liberal democracies cannot keep pace with the market's technologically driven acceleration. There is, then, a *sociology of time and power.*

A society's *future* is another socially constructed facet of its temporal system. We can conceptualize the relationship between our past and our future in terms of temporal depth. This variable concerns the distance into the past or the future that people in a given society typically imagine as they go about their daily lives. Moreover, we can hypothesize that, on average, the further backward in time a people look, the further forward they look—that is, the larger a society's past, the larger its future. This theory has practical and quite important implications for our problems with environmental hazards. As a young society with an emphasis on the short-term future, our difficulty in dealing with threats to the environment is related to the fact that they tend to develop so slowly and over such long periods of time that we have trouble perceiving them.

Variation in Our Experience of Time

Despite the standardization of time, there is *subjective* variation in our temporal experience. We may perceive time passing slowly or quickly relative to the standard temporal units of clocks and calendars. And, of course, there is the third possibility that we may experience a rough synchronicity between subjective and objective time. This form of temporality may involve subjectivity, but it is no less social in its etiology than the International Date Line. Our perception that time is passing quickly or slowly always occurs against the backdrop of standard (i.e., socially constructed) temporal units and a normal, albeit approximate, synchronicity between subjective and objective temporality. Furthermore, variation in the perception of time is conditioned by our social circumstances, and this causal impact is mediated by a thoroughly socialized entity—namely, self-consciousness. Several theories link the social structure of self-consciousness with variation in temporal experience.

In any given society, one learns to negotiate multiple forms of reality. A person in our own society, for instance, must recognize that art, religion, science, play, and dreams are different realms of meaning. They represent unique worlds of experience. Each of them is characterized by an internal consistency of meaning and experience that is divergent from what we find in the other realms. A theatrical production, novel, or film may use flashbacks to create scenes that occurred prior to the present action in the plot. Our religion may ask us to contemplate a god who exists outside time in an unfathomable eternity. Scientists examine processes that transpire in nanoseconds or billions of years. Children at play may lose all track of time as they are absorbed by a fanciful game of make-believe. Our dreams are marked by temporal experience that is wildly unfettered by any of the usual restrictions. We experience the self and time differently in these (and other) worlds, which enables us to recognize transitions from one realm of reality to another.

Another theory posits that our sense of duration results from the self's ability to integrate recollection and anticipation. According to this theory, variation in the perceived passage of time reflects the density of experience per standard temporal unit. Time is perceived to pass slowly when the density of self-conscious information processing is high. The circumstances that occasion this experience include intensely pleasant or unpleasant emotions, violence or danger, waiting and boredom, altered states of consciousness, concentration and meditation, or shock and novelty. Time is perceived to have passed quickly when the density of self-conscious information processing is low. These circumstances are brought about in two different ways. First, one is confronted by a challenging but unproblematic situation (i.e., a busy night at work). Given some familiarity with, or training for, this situation, one can act with little self-consciousness or attention to time itself, which reduces the density of experience per standard temporal unit. Second, the erosion of episodic memory lowers the density of experience in almost all remembered intervals, making for the almost universal feeling that "time flies." Time is perceived to be roughly

synchronized with clocks and calendars when the density of experience per standard temporal unit is moderate. Here, a combination of temporal socialization and routine social experience enables us to translate the flow of subjective experience into standard temporal units, and vice versa.

The foregoing theory assumes that variation in the perception of time is determined by one's circumstances. More recently, it has been theorized that temporal experience is conditioned by self-determination via "time work"—that is, calculated effort by individuals and groups to control, manipulate, and customize their experience of time or that of others. By means of folk theories and practices, we make time seem to pass more quickly or slowly than it would have otherwise (duration). We decide how often certain things should happen (frequency). We arrange our conduct in a particular order (sequence). We choose when to engage in various activities (timing). We set time aside for the people and pursuits that matter in our lives (allocation). And we take time from others, notably our employers (the theft of time). From this perspective, multiple dimensions of temporal experience are produced by our intervention or forbearance.

Michael G. Flaherty

See also Collective Memory; Self and the Social Sciences; Social Construction of Reality; Social Conventions; Time, Philosophical Theories of

Further Readings

Adam, B. (1998). *Timescapes of modernity: The environment and invisible hazards*. London, England: Routledge.

Auyero, J. (2011). Patients of the state: An ethnographic account of poor people's waiting. *Latin American Research Review, 46,* 5–29.

Bluedorn, A. C. (2002). *The human organization of time: Temporal realities and experience*. Stanford, CA: Stanford University Press.

Flaherty, M. G. (1999). *A watched pot: How we experience time*. New York, NY: New York University Press.

Flaherty, M. G. (2011). *The textures of time: Agency and temporal experience*. Philadelphia, PA: Temple University Press.

Heidegger, M. (1962). *Being and time* (J. Macquarrie & E. Robinson, Trans.). London, England: SCM Press. (Original work published 1927)

Husserl, E. (1964). *The phenomenology of internal time-consciousness* (J. S. Churchill, Trans.). Bloomington: Indiana University Press. (Original work published 1928)

Mead, G. H. (1932). *The philosophy of the present*. Chicago, IL: University of Chicago Press.

Rosa, H. (2003). Social acceleration: Ethical and political consequences of a desynchronized high-speed society. *Constellations, 10,* 3–33.

Schutz, A. (1962). *Collected papers: Vol. 1. The problem of social reality* (M. Natanson, Ed.). The Hague, Netherlands: Martinus Nijhoff. (Original work published 1945)

Schwartz, B. (1975). *Queuing and waiting: Studies in the social organization of access and delay*. Chicago, IL: University of Chicago Press.

Sorokin, P. A., & Merton, R. K. (1937). Social time: A methodological and functional analysis. *American Journal of Sociology, 42,* 615–629.

Zerubavel, E. (1982). The standardization of time: A sociohistorical perspective. *American Journal of Sociology, 88,* 1–23.

Zerubavel, E. (1985). *The seven day circle: The history and meaning of the week*. New York, NY: Free Press.

Transcendental Arguments

Transcendental arguments constitute an important type of argumentation strategy in post-Kantian philosophy. While they have mostly been employed in general epistemology, often in attempts to overcome skepticism about the external world or other minds, they are also relevant to philosophical explorations of the social sciences. This entry first sketches the structure of transcendental arguments and then provides examples of their use in the philosophy of science in general and the philosophy of the social sciences in particular.

As it will turn out, it is in some cases debatable whether a given philosophical argument or piece of reflection should be classified as "transcendental" or not. The current debate over transcendental arguments is an indication of the difficulties in strictly separating transcendental philosophies from nontranscendental ones. There may be no single essential criterion that distinguishes transcendental arguments from nontranscendental argumentation. The context is decisive.

The History and Structure of Transcendental Arguments

Although Immanuel Kant himself only rarely used the term *transcendental argument,* this type of argument is largely based on his *Critique of Pure Reason* (1781/1998). After Kant, philosophers of quite different persuasions—such as Edmund Husserl, Charles S. Peirce, and Ludwig Wittgenstein—have engaged in transcendental inquiries, though it is often controversial how exactly a certain argument available in their works should be understood. In contemporary philosophy, a new debate over the nature of transcendental arguments, which is to some extent still going on, was launched by P. F. Strawson and Barry Stroud in the 1950 and 1960s.

Generally, a transcendental argument seeks to demonstrate that something is a *necessary condition* for the *possibility* of something whose actuality is taken for granted or considered indubitable. The historical paradigm case is Kant's project of demonstrating the necessary applicability of the forms of intuition (space and time) and the pure concepts of the understanding, or the categories (e.g., causality), to all objects of humanly possible experience. Kant tried to show that certain sensible and intellectual conditions must obtain if there is any cognitive experience of a structured, nonchaotic reality. As we undeniably do have such experience, the argument concludes that the relevant conditions obtain, for instance, that all experienceable objects and events are spatiotemporal.

A transcendental argument can be schematically presented as follows:

1. If A is possible, then C.
2. A is possible (because actual).
3. Therefore, C.

As such, this schema does not differ from an ordinary *modus ponens* inference. There is, of course, an aspect of *necessity* involved here: Necessarily, if A is possible, the condition C obtains. But it is unclear whether this is anything else but the logical and/or conceptual necessity of any deductively valid argument. Clearly, necessity cannot be attached to the obtaining of C itself, because this is contingent. According to Kant himself, it is contingent that there *is* any experience; a fortiori, it is contingent that its conditions obtain, if they do. The relevant kind of necessity in a transcendental argument is *presuppositional: If* there is (or can be) experience, *then* its conditions must obtain. Typically, this presuppositional necessity is taken to be a powerful weapon against the skeptic: Even the skeptic cannot coherently doubt that there is experience; by allowing that, the skeptic must allow that experience, and its objects have a certain conceptual (categorial) structure.

Mere argument form does not provide us with the "essence" of transcendental arguments. It may be suggested that transcendental arguments—and, more generally, transcendental philosophy as philosophy employing such arguments—have a "family resemblance" character. There can be quite different uses of transcendental arguments depending on the *contexts of inquiry* in which they are set. For example, Kant's employment of transcendental reasoning is connected with his *transcendental idealism,* according to which the spatiotemporal world structured by the categories is not the world of "things in themselves" but a humanly constructed phenomenal world. (However, in the Kantian tradition, the transcendental must *not* be confused with the *transcendent:* Whereas transcendental philosophy examines the conditions and limits of experience, the transcendent—e.g., the things in themselves—lies beyond those limits. Transcendental philosophy or transcendental arguments need not deal with anything transcendent.)

In contrast to Kant's idealism, in the modern epistemological debate, Stroud argued in "Transcendental Arguments" that it remains unclear whether transcendental arguments can overcome skepticism without presupposing idealism. His criticism was directed at Strawson's use of transcendental arguments in *Individuals*. Despite their disagreement about the success of these arguments, Strawson and Stroud agreed that they should be disconnected from transcendental idealism. Others, including Kant scholars like Henry Allison, have suggested that such a connection is vital to whatever success those arguments have.

While many philosophers insist that transcendental arguments, to be interesting, should be inherently antiskeptical, others acknowledge more moderate types of transcendental argument, aiming at clarifications of our conceptual commitments instead of any knocking down of skepticism. Few philosophers maintain that transcendental arguments can reach conclusions about reality as it is, independently of the conditions of experience. Whether this means

that those arguments fall short of what they should achieve or whether this is, rather, something to be expected depends on one's overall philosophical commitments, especially regarding realism and idealism.

Transcendental Arguments in the Philosophy of Science

Philosophers of science have only rarely explicitly employed transcendental arguments, but most key positions in 20th-century philosophy of science rely on Kantian assumptions. The logical empiricists, such as Rudolf Carnap, presupposed a version of the distinction between the empirical and the transcendental in distinguishing between existence questions "internal" and "external" to linguistic frameworks. Carnap's suggestion that external questions concerning the choice of a linguistic framework are practical questions not to be decided theoretically but in terms of the fruitfulness of the framework contributed to setting transcendental philosophy (of science) on a more pragmatic path.

Thomas Kuhn's famous theory of scientific paradigms, presented in *The Structure of Scientific Revolutions,* is, again, not an explicitly transcendental position; indeed, given Kuhn's historicism, it appears to be contrary to transcendental endeavors, but Kuhn may be interpreted as claiming that, at a normal-scientific stage, a paradigm is a necessary condition for the possibility of there being scientific entities or truths about them. Scientific representations of the world would be impossible without socially shared frameworks of ontological and methodological commitments (paradigms); such representations are possible because actual; hence, science is (mostly or perhaps essentially) an activity taking place within paradigms. An additional Kuhnian insight, based on extensive empirical documentation, is that paradigms change historically. The contexts making scientific representations possible are thus not fixed once and for all. The variable C in the transcendental argument schema is, then, truly variable.

Although transcendental arguments seem to have their "timeless" deductive form, they can be operative within historically and pragmatically relativized contexts. The use of transcendental arguments need not, then, be based on ahistorical and essentialist conceptions of philosophy. This is why they can be useful also in contemporary practice–oriented philosophy of (social) science. Moreover, as transcendental arguments seem to lack any unifying essence, it may be more useful to describe philosophical approaches, methodologies, or traditions—instead of isolated arguments—as transcendental.

Transcendental Arguments in the Philosophy of the Social Sciences

The obvious place to look for transcendental arguments in the philosophy of the social sciences (broadly understood) is *hermeneutics,* which seeks to establish conditions for the possibility of meaning and/or communication. Discussions of the necessary presuppositions of communicability and intelligibility can be found in very different philosophies of meaning and understanding, including Martin Heidegger's and Hans-Georg Gadamer's hermeneutics examining understanding as a mode of being in the world, Karl-Otto Apel's and Jürgen Habermas's discourse ethics and theory of communicative action, and Donald Davidson's theory of triangulation. These and other thinkers have offered a variety of "repositionings" of the idea of the transcendental.

Just like the paradigm case of *epistemic* transcendental arguments is Kant's theory of the categories as necessary conditions for the possibility of experience—itself modified and reconceptualized in various ways both in the analytic epistemological discussions following Strawson and Stroud and in the post–logical-empiricist and post-Kuhnian developments in the philosophy of science; the paradigm case of *semantic* transcendental arguments focusing on meaning and understanding is, presumably, Ludwig Wittgenstein's *private-language argument,* presented in *Philosophical Investigations*. Wittgenstein himself never attached this label to his attempt to demonstrate that there can be no such thing as a private language—in the strict sense of a language that only the speaker or user herself could understand—and scholars have debated not only on what the private-language argument exactly seeks to show and possibly succeeds, or fails, in showing but also on whether there is such an argument in the *Investigations,* and even whether there are any traditional philosophical arguments or theses to be found in Wittgenstein at all.

The private-language argument can be seen as straightforwardly transcendental (though the scholarly controversies on this issue must be skipped here): It is, according to Wittgenstein, a necessary condition

for the possibility of meaning (or meaningful communication or meaningful use of language) that meanings are public. Meaningful communication (or use of language) is actual, hence possible; therefore, meanings are public. Thus, necessarily, insofar as there is any communication of meanings or any normatively constrained language use at all, meanings cannot be private to the speaker. Language need not be essentially social, as we may imagine Robinson Crusoe using a solitary language in isolation, but it must be public in the sense that its meanings are in principle open to others. It must also be normatively structured. Otherwise, it is no language at all.

Arguments of this kind, referring to the possibility of meaning and communication, are highly relevant in the philosophy of the social sciences. Following Wittgenstein, Peter Winch challenged "positivistic" social sciences by arguing that social practices—or what Wittgenstein called "forms of life"—should not be causally explained but should be internally understood by learning to understand their rules, just as we learn to play a "language-game" by learning to follow its rules. While many philosophers maintain that Winch went too far in embracing *relativism,* the view that social practices (or cultures, traditions, perspectives, etc.) are only intelligible and rationally discussable in their own terms, "from within"—a view that seems to make any critical dialogue across practices impossible—the transcendental character of Winch's and other Wittgensteinians' arguments can be appreciated without drawing their extreme conclusions. To engage in causally explanatory social science, one must understand the phenomena to be explained, and transcendental arguments and inquiries may play a significant role in clarifying the conceptual networks underlying those phenomena, as well as our conceptual commitments and "pre-understandings."

Transcendental arguments are, thus, also connected with the *Erklären* versus *Verstehen* ("explanation vs. understanding") debate over the methodology of the social sciences, a debate that may have lost its urgency since the 1960s and 1970s but may still characterize important divergences among social-scientific approaches. There is no consensus about the philosophical relevance of transcendental arguments, or even about their nature, but they continue to attract not only scholars interpreting Kant or Wittgenstein but also those systematically working on issues in knowledge and meaning.

Sami Pihlström

See also A Priori and A Posteriori; Analytic/Synthetic Distinction; Epistemology; Explanation Versus Understanding; Idealism; Language-Games and Forms of Life; Pragmatism; Relativism and the Social Sciences: From the Sapir-Whorf Hypothesis to Peter Winch; Transcendental Pragmatics

Further Readings

Allison, H. E. (2004). *Kant's transcendental idealism: Interpretation and defense* (Rev. ed.). New Haven, CT: Yale University Press.

Carr, D. (1999). *The paradox of subjectivity: The self in the transcendental tradition.* Oxford, England: Oxford University Press.

Kant, I. (1998). *Critique of pure reason* (A. Wood & P. Guyer, Trans.). Cambridge, England: Cambridge University Press. (Original work published 1781)

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago, IL: University of Chicago Press.

Malpas, J. (Ed.). (2003). *From Kant to Davidson: Philosophy and the idea of the transcendental.* London, England: Routledge.

Pihlström, S. (2004). Recent reinterpretations of the transcendental. *Inquiry, 47,* 289–314.

Stern, R. (Ed.). (1999). *Transcendental arguments: Problems and prospects.* Oxford, England: Clarendon Press.

Strawson, P. F. (1993). *Individuals: An essay on descriptive metaphysics.* London, England: Routledge. (Original work published 1959)

Stroud, B. (2000). Transcendental arguments. In *Understanding human knowledge* (pp. 9–25). Oxford, England: Oxford University Press. (Reprinted from "Transcendental Arguments," by B. Stroud, 1968, *Journal of Philosophy, 65,* pp. 241–256)

Winch, P. (1958). *The idea of a social science and its relation to philosophy.* London, England: Routledge & Kegan Paul.

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Oxford, England: Blackwell.

Transcendental Pragmatics

This entry introduces Karl-Otto Apel's philosophical thought, which has been of interest to the philosophy of the social sciences. It explains the central feature of Apel's theory of the community of communication, and the nature of pragmatic transcendental conditions (hence a priori) necessary for meaningful discourse, and ends by indicating its

implications also for ethics. Rather than philosophy as a theorizing *in abtracto,* Apel's theory places the philosophical quest for truth and other such philosophical aims at the center of an already existing linguistic community within which theorizing is always already set—thus emphasizing the nonsolipsistic foundations of critical philosophical exercise.

Introduction

"Transcendental pragmatics" is the name given by the German philosopher Karl-Otto Apel (b. 1922, Düsseldorf, Germany) to his philosophical program of a linguistic transformation of transcendental philosophy. Inspired by Martin Heidegger, Hans-George Gadamer, Ludwig Wittgenstein, and above all Charles Sanders Peirce, Apel fused what he called "transcendental hermeneutics" with "transcendental semiotics" in order to articulate the project of critical reflection on the conditions of possibility of valid intersubjective communication.

Apel first developed this philosophical program in a series of essays written during the late 1950s and 1960s. The essays were collected in a two-volume work titled *Transformation der Philosophie,* published in 1973, which was only partly translated into English in 1980. The basic claim is that anyone who seeks truth—whether as philosophers, scientists, or humanists—has always already entered an argumentative discourse that presupposes a series of conditions of possibility. More specifically and succinctly, anyone who makes any claim to truth does so as a member of a community of communication.

The A Priori of the Community of Communication

Using Wittgenstein's argument against the impossibility of a "private language," which claims that no individual is able to have a private, or solely mental, language that he or she alone can speak, Apel arrives at the insight that there is no individual who is not always already a member of a community of communication. Humans are linguistic beings, but only as members of a specific linguistic community, who thus find themselves immersed in historical worlds whose pre-understandings, or ways of giving meaning, condition how they understand the world, others, and themselves. Yet although our community of communication is always a community of a specific natural and historical language, Apel argues that we are not therefore caught in the grip of an inescapable relativism. All communication, regardless of the language in which it is undertaken, presupposes that we aim to reach an agreement about something with someone. All languages have a propositional-performative structure, or dual structure. When we speak, we make claims about the world, a state of affairs, or something, in order to come to an agreement or understanding with someone else.

This dual structure of all languages means that whenever we engage in some argumentative discourse we presuppose four validity claims: (1) a claim to communicable meaning or sense, (2) a claim to truth, (3) a claim to veracity or truthfulness, and (4) a claim of rightness or correctness. When we engage in any kind of communication or discourse, we inescapably and inevitably presuppose that we are making sense, or that we are being intelligible; that we also mean what we claim, or rather that what we claim is something we believe; that what we claim is a truth claim that can be verified; and, finally, that when we claim some truth about a state of affairs or fact in the world, we do so to someone with whom we have thus established a right relation as an equal partner in a discourse.

Performative Self-Contradiction and Validity Claims

That all discourse presupposes these validity claims is far more intuitive than it may appear prima facie. Some counterexamples demonstrate the "transcendental" character of these presuppositions—that is, that without them all discourse, all meaningful argumentation, becomes impossible or meaningless. One counterexample could be as follows: I claim *x,* but I don't believe what I am claiming; *or* I claim *x,* but I am doing something in a nonsensical way or a way that only I can understand (via some secret code that I alone can crack); *or* I claim *x,* but there is no way anyone can verify it; *or* I claim *x,* but you are either too stupid or, because of your race or gender or religion, incapable of understanding or verifying it. To deny or pretend that any of these validity claims can be bracketed, neutralized, or suspended leads to a performative self-contradiction: that is, I am denying what I am doing. Reflecting on what conditions of meaningful communication cannot be denied or refuted without incurring a performative self-contradiction reveals what conditions of possibility of valid thought or meaningful discourse are

"uncircumventable" (*nichthintergehbar*). The performative self-contradiction becomes the litmus test that reveals to us the *pragmatic transcendental* conditions of possibility of discourse *tout court*. These conditions are *pragmatic* because they are indispensable to all and every act of communication. With every speech act we enunciate, we are simultaneously saying and doing. The conditions of possibility of meaningful communication are transcendental and pragmatic in the sense that they are a priori; they enable and condition meaningfulness, but they are always already operative in the very performance of every speech act.

Ultimate Foundation, Self-Recuperative Principle, and Discourse Ethics

Apel's philosophical project, however, is not simply a linguistic transformation of transcendental philosophy that aims to preserve the Kantian project of transcendental reflection on the conditions of possibility of valid cognition, but now as a transcendental-pragmatic reflection on the conditions of meaningful discourse, it is also a comprehensive philosophical proposal with several pillars. These pillars also reveal how Apel's project is different from Jürgen Habermas's "universal or formal pragmatics," with which it shares some features. Transcendental pragmatic aims to be a synthesis of both transcendental hermeneutics and transcendental semiotics. The hermeneutical investigation into the conditions of valid understanding, which is not simply about understanding differently but understanding better, is fused with Peircian semiotics (from the American pragmatist Charles Sanders Peirce), whose basic insight is that all cognition, and discourse, is mediated by indexes, icons, and symbols. All understanding takes place in and through semiosis.

Apel positions his transcendental pragmatics as a "first philosophy," *prima philosophia,* which supersedes the prior paradigms of first philosophy, namely, ontological metaphysics (from Aristotle through Aquinas), and transcendental philosophy of consciousness (from René Descartes to Edmund Husserl). As a *prima philosophia,* transcendental pragmatics offers a non-metaphysical and nonsolipsistic ultimate foundation (*Letztbegründung*) that reveals the conditions of possibility of all critique and reflection, thus refuting all relativism, hyperbolic methodical doubt, and incommensurability of hermeneutical horizons. Most interestingly, Apel's philosophical project also offers a normative reconstruction of philosophy from the standpoint of the semiosis of all cognition and understanding as a reconstruction of the ways in which different paradigms have failed to properly address the semantic, syntactical, and pragmatic dimensions of sign use. Apel argues, furthermore, that his project includes the elaboration of what he called the "self-recuperative principle" (*Selbsteinholungsprinzip*), which holds that human history, and thus the human and natural sciences, can be reconstructed from the quasi-teleological standpoint of a better understanding that can give an account of itself in a normative and noncontingent way. All understanding and critical discourse presuppose progress, one that can be elucidated by reflection on what enables our present ability to come to a valid intersubjective agreement. Finally, transcendental pragmatics also elucidates the foundations of a discourse ethics or normative morality, which in turn entails an *ethics of responsibility.* Transcendental pragmatic reflection on the uncircumventable conditions of possibility of valid intersubjective communication reveals the always already implicit recognition of some basic moral norms: isonomia and reciprocal recourse to dialogue in order to come to an agreement about what principles should guide our ethical action.

Eduardo Mendieta

See also Communicative Action Theory; Frankfurt School and Critical Social Theory; Language-Games and Forms of Life; Pragmatism; Semantics and Pragmatics; Transcendental Arguments

Further Readings

Apel, K.-O. (1980). *Towards a transformation of philosophy.* London, England: Routledge & Kegan Paul.

Apel, K.-O. (1993–1996). *Selected essays* (2 vols.). Atlantic Highlands, NJ: Humanities Press.

Apel, K.-O. (1998). *From a transcendental-semiotic point of view.* Manchester, England: Manchester University Press.

Mendieta, E. (2002). *The adventures of transcendental philosophy: Karl-Otto Apel's semiotics and discourse ethics.* Lanham, MD: Rowman & Littlefield.

Transhumanism and Human Enhancement

Human enhancement encompasses efforts to extend human health, longevity, and cognitive abilities beyond their current biological limits. *Transhumanism* is the Enlightenment-inspired doctrine that science can accomplish human enhancement and that individuals should be able to use technologies that allow them to live longer, healthier, and more enabled lives. The emergence of a global, self-consciously transhumanist movement in the late 20th century has been accompanied by growing diversity within transhumanism and a growing number of bioconservative critics from the traditional Left and Right.

Proto-Transhumanism and Enlightenment Bio-Utopianism

The aspiration to transcend human limitations is found in the earliest recorded human cultures. The Epic of Gilgamesh, for instance, is the story of a man searching for immortality. Shamanic and religious traditions attempted to use supernatural means to heal and achieve superhuman powers and immortality. But the European humanist and Enlightenment traditions established a new direction for these aspirations, combining faith in reason and individual self-governance with utopian expectations of a free and bountiful technological future. For Enlightenment materialists like Julien de La Mettrie, human beings are not confined to their bodies and brains by divine will but by chance and have the power to become something better. The Marquis de Condorcet, Benjamin Franklin, and William Godwin proposed that eventually human beings would be able to conquer death, and Denis Diderot suggested that humanity might evolve into a great variety of *post-human* species.

Enlightenment thought also began to interrogate premodern ideas about human uniqueness based on divine creation and ensoulment and proposed that rational and moral subjectivity were the basis of moral and political standing. If rational and moral faculties were not supernatural and could be found in animals, as proposed by David Hume, for instance, then radically evolved humans, animals, and even machines might be rational, emotional, and moral subjects. Denis Diderot proposed in *D'Almbert's Dream* that brains might be taken apart and reconstituted later, that intelligent animals and animal–human hybrids might be possible, and that sophisticated machines might have minds.

Many of the debates among contemporary transhumanists, and between them and their secular critics, can be seen as legacies of the contradictions between different strains of Enlightenment thought.

Nineteenth- and Early-Twentieth-Century Transhumanism

The improvement of the human condition through social reform took precedence over the *bio-utopian* imagination in the 19th century, until the emergence of eugenics. The eugenicists believed that the future of humanity could be improved by discouraging childbearing in groups with bad heritable traits ("idiots," criminals, the poor in general) and encouraging reproduction by those with better traits. Some have argued that transhumanism is a modern form of *eugenicism,* albeit a liberal one that proposes genetic betterment through individual germinal choice and gene therapy rather than mandated sterilization, abortion, and murder.

Transhumanists, on the other hand, identify with bio-utopians like the British Marxist geneticist J. B. S. Haldane, who rejected the pseudoscience and authoritarianism of eugenics and proposed instead, in his 1923 seminal essay *Daedalus, or a Science and the Future,* that people would be able to choose their own genetic traits in the future. Haldane's friend and fellow-geneticist Julian Huxley in 1927 coined the term *transhumanism* to encompass the belief that humanity could, scientifically and spiritually, transcend itself.

Speculative fiction also began to explore *biofuturism*. The Fabian socialist H. G. Wells alternated between the dystopian biofuture of *The Time Machine* and the optimistic utopianism of *Men Like Gods* (1923) and *The Open Conspiracy* (1928). Olaf Stapledon inspired a generation of writers with the breadth of post-human options presented in his 1930 *Last and First Men* and 1937 *Star Maker.* Julian's brother Aldous Huxley wrote *Brave New World* in 1932, partly as a critique of the bio-utopian ideas in their British social circle.

In 1926, the Irish Marxist and scientist J. D. Bernal contributed another strain to contemporary transhumanism with his essay *The World, the Flesh and the Devil.* Bernal proposed that humans would eventually colonize space in genetically modified *cyborg* bodies with brains linked to machines.

After the defeat of fascism and the widespread rejection of anything associated with eugenics, combined with the growth of nuclear anxiety and ecological awareness and the emergence of antirationalism and pastoralism in the counterculture, bio-utopianism nearly disappeared in the 1950s.

Counterculture Meets High Technology

In the 1960s, however, numerous trends began to reignite the bio-utopian imagination. One was the emergence of an antiaging subculture, the most radical exponents of which were the cryonicists. In his 1962 *The Prospect of Immortality* and 1972 *Man Into Superman,* cryonics pioneer Robert Ettinger proposed that the human body could be radically redesigned to be immortal, to fly, to swim like a fish, to photosynthesize, and to eventually become a galaxy-spanning brain. Techno-optimistic futurists, developing out of corporate and national security consulting, began discussing the ramifications of trends like artificial reproductive technologies and brain–machine interfaces. For instance, the New York City–based futurist "FM-2030" (born Fereidoun M. Esfandiary in Iran) began discussing our period of history as "transhuman," transitional to the post-human, and promoted putatively transhuman lifestyles and value systems in his 1970 *Optimism One* and 1973 *Upwingers: A Futurist Manifesto*. The rapidly growing science fiction subculture began to produce optimistic visions of a post-human future in works by writers such as Cordwainer Smith, Robert Heinlein, and Frederick Pohl. A small group of feminists, such as Shulamith Firestone and Marge Piercy, proposed that artificial wombs would liberate women from patriarchy. Responding to rapid advances in medicine, bioethicists such as Joseph Fletcher began to defend the benefits of genetic and cognitive enhancement technologies.

These trends converged in Southern California in the late 1980s around a group of futurist thinkers led by the philosopher Max More (born Max O'Connor). More founded the *Extropy Institute* and its journal, which quickly spread to an international virtual community though e-mail and the Internet. The Extropians defined transhumanism as a class of philosophies that seek to guide us toward a post-human condition, of which extropianism was the flavor aligned with anarcho-capitalism. The Extropians were especially enthusiastic about the prospect that in future nanotechnology, molecule-scaled machines would enable indefinite longevity and the uploading of consciousness to nanomachine bodies.

In the late 1990s, European transhumanists began to coalesce around the more politically inclusive *World Transhumanist Association* (WTA), founded by the Oxford philosopher Nick Bostrom and British utilitarian thinker David Pearce. In the 2000s, the WTA grew quickly, with chapters and allied groups in dozens of countries. In 2009, the WTA rebranded itself as *Humanity+*.

Transhumanist Subcultures

As the transhumanist subculture has grown, many subgroups have emerged. One division has been between the antistatist libertarians and the left-leaning transhumanists, or "technoprogressives." The Institute for Ethics and Emerging Technologies, founded in 2005 by Nick Bostrom and James Hughes, is the principal organization of techno-progressive-leaning intellectuals. These camps fall out over whether government-funded research and health and safety regulations are necessary for the development of emerging technologies and whether problems of equitable access should be addressed through the provision of universal health care.

Although most transhumanists are secular, and a sizeable group is militantly atheist, there are also many who hold a variety of spiritual views from the idiosyncratic to the orthodox. One of the largest transhumanist groups is the Mormon Transhumanist Association, for instance, which sees transhumanism as the fulfillment of Mormon prophecy.

In recent years, the millennialist subculture within transhumanism, *singularitarianism,* has also grown rapidly. The concept of "singularity" was proposed by the mathematician and science fiction author Vernor Vinge in the early 1990s as the point at which greater-than-human machine intelligence begins rapidly improving itself, effectively ending human-directed history. Most Singularitarians believe this point will occur in the 21st century. Some, such as the inventor and futurist Ray Kurzweil, believe that this

"intelligence explosion" will enable radical longevity, cure social problems like hunger and climate change, and create a new, superconnected post-human civilization. Others believe that a "Terminator-like" scenario of runaway robotics is more likely to be deflected only by determined efforts to ensure "AI friendliness." Most Singularitarians are skeptical that the transhumanist program of human enhancement and augmentation could allow human beings to stay in control of machine intelligence, given the limitations of organic brains compared with the exponential improvements in computing power.

Bioconservatives and Academic Post-Humanism

As transhumanism has grown, so also have a set of explicitly anti-transhumanist or "bioconservative" groups. These groups are motivated by a wide variety of criticisms of transhumanism, from ideas about human exceptionalism and the natural order to concerns about the safety of new technologies, and equality and quality of life in a transhuman society. Religious conservatives have started organizations like the Center for Bioethics and Human Dignity to argue against transhumanism as a form of spiritual heresy. Groups on the Left, such as the Center for Genetics and Society, argue that transhumanism is a stalking horse for racism, corporate control, and neo-eugenics. Left-leaning bioethicists George Annas and Lori Andrews mounted campaigns to make human genetic enhancement an international "crime against humanity."

The most visible critics of transhumanism have been Francis Fukuyama, whose 2002 *Our Posthuman Future* called for global regulation of transhuman technology, and the bioethicist Leon Kass, who led the U.S. President's Council on Bioethics when it published the anti-enhancement volume *Beyond Therapy* in 2003. Fukuyama's complaint focused on the erosion of a shared human identity that supposedly undergirds the political order, while Kass launched a Neo-Aristotelian critique of the effects of life extension and enhancement on human dignity and virtue. Also in 2003, the environmentalist Bill McKibben published his anti-transhumanist book *Enough,* and the German social theorist Jürgen Habermas published his bioconservative tract *The Future of Human Nature*. In 2010, the New Zealand philosopher Nicholas Agar published *Humanity's End,* the most sophisticated critique of transhumanist ideas to date.

As bioethicists polarized over cognitive enhancement in the past decade, many liberals have sided with transhumanism, for example, Greg Stock, Gregory Pence, Allen Buchanan, Art Caplan, John Harris, and Julian Savulescu. Some have gone one step further than the transhumanists to argue for a moral *obligation* for certain kinds of enhancements (e.g., especially when the "natural lottery" has disadvantaged some by birth, who need to be enhanced if the liberal principle of equal opportunities for all is to be taken seriously in practice).

A diverse group of cultural theorists have also been gathered under the banner of post-humanism, from the "cyborgologists" Donna Haraway and Chris Hables Gray, to theorists of the eroding virtualized self, like Katherine Hayles and Robert Pepperell, to postmodern critics of Enlightenment humanism, like Cary Wolfe. Some of these writers share with transhumanism a focus on cyborgs and the augmented body in reality and fiction, but they range from ambivalent to hostile to transhumanism's normative claims.

James Hughes

See also Enlightenment, Critique of; Ethical Impact of Genetic Research; Eugenics, Old and Neoliberal Theories of; Technoscience and Society; Utopianism

Further Readings

Agar, N. (2010). *Humanity's end: Why we should reject radical enhancement*. Cambridge: MIT Press.

Bernal, J. D. (1969). *The world, the flesh and the devil; an enquiry into the future of the three enemies of the rational soul*. Bloomington: Indiana University Press.

Bostrom, N. (2003). *The transhumanist FAQ* (v. 2.1). Retrieved from http://www.transhumanism.org/resources/FAQv21.pdf

Bostrom, N. (2005). A history of transhumanist thought. *Journal of Evolution and Technology, 14*(1), 1–30.

Condorcet, J. (1979). *Sketch for a historical picture of the progress of the human mind*. Westport, CT: Greenwood.

Ettinger, R. (1972). *Man into superman*. New York, NY: St. Martin's Press.

Firestone, S. (1970). *The dialectic of sex: The case for feminist revolution*. New York, NY: Morrow.

Fukuyama, F. (2002). *Our posthuman future: Consequences of the biotechnology revolution*. New York, NY: Farrar, Straus & Giroux.

Garreau, J. (2006). *Radical evolution: The promise and peril of enhancing our minds, our bodies—and what it means to be human*. New York, NY: Broadway.
Gray, C. (2000). *Cyborg citizen: Politics in the posthuman age*. New York, NY: Routledge.
Gray, C. (Ed.). (1995). *The cyborg handbook*. New York, NY: Routledge.
Haldane, J. (1924). *Daedalus or science and the future*. London, England: Kegan Paul, Trench, Trubner.
Hughes, J. (2004). *Citizen cyborg: Why democratic societies must respond to the redesigned human of the future*. Cambridge, MA: Westview Press.
Kaldis, B. (2011). Brave new world: Transhumanism, prosthesis and enhancement debates. In F. Nadis & J. Waskey (Eds.), *ABC CLIO world history encyclopedia, ERA9: Promises and paradoxes* (1945–Present; pp. 771–772.). Santa Barbara, CA: ABC-CLIO.
Kass, L. (Ed). (2003). *Beyond therapy: Biotechnology and the pursuit of happiness*. New York, NY: Harper Perennial.
Kurzweil, R. (1999). *The singularity is near.* New York, NY: Viking Press.
More, M. (1990). Transhumanism: Toward a futurist philosophy. *Extropy, 6,* 6–11.
More, M. (2003). *Principles of extropy* (Version 3.1). Retrieved from http://www.extropy.org/principles.htm
Regis, E. (1990). *Great mambo chicken and the transhuman condition*. Reading, MA: Addisonh-Wesley.
Savulescu, J., & Bostrom, N. (Eds.). (2011). *Human enhancement*. Oxford, England: Oxford University Press.

Trust, Epistemic

Knowledge is a common good. Relying on others in acquiring knowledge is part of our ordinary cognitive life. The floating of other people's words in our minds is the price we pay for thinking. This doesn't concern ordinary beliefs only: Collaborative work in contemporary science is such that scientists must trust each other in order to achieve a relevant epistemic result. Trust is thus a fundamental ingredient not only of our social life but also of our epistemic practices.

The study of epistemic trust in philosophy and in social science encompasses a wide range of questions about the role of trust in knowledge, the division of cognitive labor in society, collaborative work in science, and deference to the authority of experts. This entry reviews the enhanced role given to epistemic trust in recent accounts of epistemology, ethics, and the social sciences.

The Received View: The Autonomous Knower

Traditionally, epistemology has banned from actual knowledge beliefs acquired by trusting others. One of the strongest requirements for the acquisition of knowledge is the *autonomy* of the subject. The overall project of classical epistemology, from Plato to contemporary rationalist approaches, is a normative enterprise aiming at establishing criteria, rules, and principles as a guarantee to preserve the autonomy and freedom of thought necessary to the acquisition of knowledge. Authority, received opinions, and common knowledge are considered in this tradition as the major sources of false beliefs. To strengthen the cognitive autonomy of the rational thinker, philosophers of all times have listed rules of "epistemic conduct" that guarantee freedom of thought. Rene Descartes's classic treatise on method, *Rules for the Direction of Natural Intelligence,* is dedicated to explaining how people should think in order to attain true ideas and bases its model of the autonomous knower on self-reflection or self-contained meditation. In the empiricist tradition, John Locke, in his *Essay Concerning Human Understanding,* insists on the risk of being "infected" by other people's opinions and lists a series of obligations on one's own mental conduct to avoid contamination by ill-formed beliefs.

Challenging the Traditional View

Yet the massive reliance on others that permeates our cognitive life calls for an epistemic treatment. This has become a central issue in contemporary debates in philosophy of knowledge and *social epistemology.* A number of approaches have been put forward to account for the epistemic reliability of the division of cognitive labor so typical in contemporary, information-dense societies. The received image taken from the traditional epistemology of the solitary scientist, the autonomous knower who has absolute control of the sources of his or her beliefs, is a remote ideal, a limit case that doesn't correspond to the reality of our epistemic practices. One could even argue that it is not even an ideal: A distribution of epistemic competencies makes a society more efficient and rational than does a concentration of knowledge within a tiny group of experts.

Trust in Social Sciences and Moral Philosophy

Trust is a central notion of social science. It is considered as the "glue" of our society, the mechanism that is at the basis of social relations, of transfer of power to political authority, and of cooperation. Various approaches in the social sciences have tried to account for the notion of trust. In the *rational choice* tradition, trust is seen as a cognitive capacity. The sociologist Diego Gambetta (1988) defines it as

> a level of subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both *before* he can monitor such action (or independently of his capacity ever to be able to monitor it) *and* in a context in which it affects *his own* action. (p. 218)

The political theorist Russell Hardin defines trust as a form of *encapsulated interest,* that is, a belief that it is in the interest of the trusted to attend to the trustor's interests in the relevant matter. These are *evidential* approaches to trust that see it as knowledge or belief for which we can find a rational justification in terms of the capacity we have to read and assess the commitments of others. Other approaches, which one may call *motivational,* see trust not only as a cognitive competence, that is, based on the degree of our beliefs about the future actions of the trusted, but also as involving a motivational, nonrepresentational dimension that may depend on our deep moral, emotional, or cultural precommitments.

These approaches have been developed mainly in sociology and moral philosophy. The philosopher Annette Baier defines trust as an accepted vulnerability to another's possible but not expected ill will toward one and explores the varieties of moral, emotional, and cultural grounds on which we accept this vulnerability.

Trust in Epistemology

One of the main tasks of the investigation of epistemic trust is to understand which concept of trust applies in the case of knowledge acquisition. Is trust in other people's beliefs a cognitive capacity to assess the probability of their reliability? Is it an accepted vulnerability based on forms of precommitments or deference to authority? In which cases are we *justified* in trusting what other people say? These questions evoke another important contemporary debate in social epistemology, that is, the status of *testimonial knowledge* and the *epistemology of testimony.* Reductionist approaches to testimony state that justification of testimony is always reducible to more basic forms of knowledge acquisition, such as perception and inference. Nonreductionist approaches claim that we have an a priori justification in trusting other people's beliefs, in the absence of any rational *defeaters* of that belief, that is, any stronger evidence that the testimony we receive is false.

The main criticism of the application of rational-choice approaches to trust in the case of epistemology revolves around the fact that, in the case of knowledge acquisition, we should be able to assess not only the willingness of our informants to be trustworthy but also their competence. A 5-year-old child can be trustworthy but lacking the appropriate competence to be a reliable informant. Also, in most cases of belief acquisition by trusting others, we do not have the means to estimate the subjective probability of the informant to be right. A patient who trusts her doctor and a child who trusts her parents are not able to calculate the odds at stake. Motivational approaches to trust seem to deal better with this problem, by providing reasons to trust that go beyond pure assessment of the evidence available. But they also bear problems if applied to the epistemic case, because they seem unable to avoid the risk of credulity and irrationality that accompanies prima facie any a priori trust in others as a source of knowledge. To understand what motivates us to trust others in acquiring knowledge, we should be clear about the precommitments that sustain our cognitive relations in society.

Gloria Origgi

See also Common Knowledge; Epistemology; Rational Choice and Political Science; Social Epistemology; Trust, Social; Virtue Epistemology

Further Readings

Baier, A. (1986). Trust and anti-trust. *Ethics, 96*(2), 231–260.

Daukas, N. (2006). Epistemic trust and social location. *Episteme, 3*(1–2), 109–124.

Faulkner, P. (2008). *Knowing on trust.* Oxford, England: Oxford University Press.

Gambetta, D. (Ed.). (1988). *Trust: Making and breaking cooperative relations.* Oxford, England: Blackwell.

Hardin, R. (2002). *Trust and trustworthiness.* New York, NY: Russell Sage Foundation.

Lackey, J., & Sosa, E. (Eds.). (2006). *The epistemology of testimony.* Oxford, England: Oxford University Press.
Origgi, G. (2004). Is trust an epistemological notion? *Episteme, 1*(1), 1–12.

Trust, Social

The notion of social trust as a topic of analysis and investigation is important in philosophy and in the social sciences, as it plays a crucial role in social life and especially in cooperative activities. The whole sphere of sociality rests on people being together. Without trust, people would have to rely on rules and sanctions alone. Understanding what social trust is exactly is the starting point for understanding what it takes to be trusted by, and to trust, others.

This entry focuses on an analysis of social trust that distinguishes two main types of trust from each other: *social normative* (or genuine) trust and *predictive* trust.

A rule of thumb is that when trust is betrayed the trustor had genuine trust—her trust was justified, but the trustee did not respect her rights, although he had led her to believe so. They had a relationship of mutual respect for specific rights. However, when the trustor is mistaken in her trust for the reason that her predictions were poor, she is to blame. She had predictive trust, based on some beliefs about the trustee. In genuine trust, the betrayed trustor rightfully blames the trustee, unless the trustor was mistaken about there being a mutual understanding about respect for certain rights. In predictive trust, the disappointed trustor makes the wrong predictions and has herself to blame. Of course, she could accuse the trustee of deception if he deliberately gave her the false impression. *Misplaced trust* is an ambiguous term as it can mean betrayed justified genuine trust, mistaken unjustified genuine trust, or mistaken predictive trust.

A Conceptual Analysis of Social Trust

Social trust is trust between people or groups that can function as agents. Trust is an attitude (and/or feeling) that people have when they feel at peace and have an accepting attitude about being dependent on another person for an action. The trustee should be aware of the dependence and have a choice to either accommodate the trustor or let him down. The trustor may have no other option than to trust, but he may still have genuine trust in the other person. In everyday language, to "trust someone for buying the tickets" may mean to have social normative (genuine) trust or predictive trust in her, or it may mean to (decide to) depend/rely on her, based on trust or prediction. In everyday speech, the distinction between the meanings of words often becomes blurred. To depend/rely on someone for something or to decide to do so based on a prediction of her future favorable behavior or on trust of that person should be distinguished from trusting in the sense of having an attitude or feeling of trust. "Entrusting someone with buying the tickets and after that depending/relying on her" does not, but it may entail trust. The person is made aware of what is expected of her and that people depend on her. Entrusting involves a decision, just like depending and relying usually do.

The notion of trust can be divided into two kinds depending on which of the two main kinds of mental states it is related to, respectively. These will be called (1) *social normative* (genuine) trust, based on a belief of a relationship of mutual respect for rights, and (2) *predictive* trust, based on a belief of having sufficient reasons for trusting. Genuine trust may be betrayed by the trustee, while predictive trust cannot be betrayed—it is the trustor who made the wrong prediction, leading her into having a trusting attitude and perhaps into the state of relying on another person or deciding to rely on him. In the context of economics, this latter kind of "trust" is further stretched into a mere prediction. The trustor calculates the probability for an action to take place, "decides to trust," and has various degrees of trust. The trustee might not even be aware that someone is counting on his action. In predictive trust, the attitude of trust is taken without a decision, and the trustor either trusts or not as a consequence of her calculations.

The following example suits both genuine and predictive trust. When John finds himself relying on Mary or decides to rely on her in acting on a matter of his concern, without fear of being disappointed, he may be said to trust Mary and rely on her vis-à-vis the action. He believes that Mary is aware of his dependence and that she has a choice to let him down. John may have no options but still trusts her. Alternatively, he may just trust her for potentially acting on the matter and still decide not to rely on her but choose to act himself. Predictive trust is

an attitude that John arrives at after believing that Mary can and will perform the action in a way that furthers John's welfare. He thus has an accepting attitude of being dependent on Mary and has a positive feeling about it. This takes place in a context where John wants Mary to perform the action; he believes she knows about this and that his want is relevant to her—she is free to and will take it into account in her acting. John's predictive trust is based on his reason-based beliefs, and if Mary does not deliver, John blames himself for his mistaken beliefs.

Social normative or genuine trust not only presupposes that the trustor is at peace with being dependent on the trustee, including all the beliefs of predictive trust, but it also involves the trustor's belief of a mutually acknowledged right to be accommodated in a specific case. In the above example, John's belief in such an acknowledged right is not the reason for his expectation to be accommodated by Mary, but this is the reason why he expects this *of* Mary, and this is why he feels betrayed in his trust if Mary does not perform the action. More specifically, the belief in upholding a right goes as follows: John believes that Mary and he have a relationship of mutual respect for certain rights, a belief that is based on his having certain rights and her having certain (other) rights that are respected. If the present expected action is not among the actions that he has a right to expect, he is not justified in having genuine trust in Mary. When John's genuine trust is justified, and he is not mistaken about what his respected rights are, Mary betrays his trust if she does not accommodate him. The trustor's belief of the relationship of mutual respect for rights is not just another reason for predictive trust. It is why he trusts (a causal reason)—the backbone of genuine trust.

When people have predictive trust in each other, they just count on each other to behave in the desired way, having good reasons to expect this. For example, Mary trusts that John will buy her a necklace, because she believes he is in love. She trusts the priest to give her wallet back, because she believes he is honest. When people have reasons to believe that they have rights, but they lack the belief that these are acknowledged in a relationship of mutual respect for rights, they may do no more than count on the belief that the trustee will act in accordance with those rights, on the basis of his moral character or some expected sanctions. Mary has predictive trust in John concerning his nonviolent behavior, by believing that he follows the law. Later, she has genuine trust in him vis-à-vis this matter, when she believes that they have a relationship of mutual respect for certain rights, including the right to be treated without violence.

Genuine trust is like a gift that has to be accepted by the trustee. It is an honor to be offered someone's trust, but as the gift of trust involves a demand on the trustee, he has to accept this demand. If Mary trusts John to come home early, she may have predictive trust in him, based on good reasons. For her to have genuine trust in John, she must believe that they have a relationship of mutual respect for rights. Only if it is her right (and she rightfully believes that he has accepted this), can she blame him for betrayal if he comes late. If trust does not involve the option of betrayal, it is not genuine trust. Of course, Mary may be wrong in her beliefs (based on her subjective experiences) about their mutual understanding of their charter of rights. In that case, she has misplaced genuine trust in the sense of unjustified genuine trust. Only justified genuine trust may be betrayed.

Genuine trust is not restricted to personal relationships. Two businessmen may have predictive trust in each other that will develop into genuine trust through time. When people respect each other's rights for instrumental reasons, one could call it *thin* genuine trust. In *thick* genuine trust, a person's rights are respected for intrinsic reasons: He ought to have what is his right to have. Basic trust is a child's default position when cared for. Some adults preserve their basic trust for people to a larger extent than others. They may fare poorly in bad company, but on the other hand, a trusting attitude invites people to be fair. General trust in a person is trust in his or her fairness concerning a wide range of matters—that is, rightful demands will be accommodated. People may have general predictive trust in a member of the clergy, based on his position, and general genuine trust in a friend, based on a relationship of mutual respect for a wide range of rights.

The relationship of mutual respect for rights in the case of genuine trust could in extreme cases be a relationship of one-sided respect for rights, such as in a situation where one party is in no position to accommodate the other party. For example, a severely handicapped wife cannot be expected to respect in practice her husband's rights in any way. Outside such extreme examples, mutuality is needed

for stability of the trust relationship. One could even argue that the golden rule would be a central principle for the charter of rights in a relationship of mutual respect for rights in social normative trust.

Trust has been studied as a commitment, a relationship, an attitude, an emotion, or a mixed case, for example, a belief with an affective component. Predictive reliance on and prediction of a person's behavior allow for depending on the side effects of his action. Trusting requires a context where the trustee is believed to be aware of the trustor's wish to be (potentially) dependent on the trustee's action and to be free to choose to take the trustor's wish into account and free to let him down. The belief of a person's trustworthiness is a result of predictive trust or genuine trust or a judgment preceding predictive trust. In the context of prediction, it is the result of an evaluation before the decision to depend on the person. Trusting involves risks, but only when seen from a third-person point of view. When a person trusts, he may be aware of the third-person point of view, but he does not deem it risky, or else he would not trust. This holds for both predictive trust and social normative (genuine) trust—the two main kinds of trust that one does not decide to have but one just "falls into," like falling in love. Risk calculations belong to predictions, and so do degrees of "trust." People may decide to depend/rely on others on the basis of their predictions.

Maj Tuomela

See also Action, Philosophical Theory of; Commitment; Cooperation/Coordination; Normativity; Social Capital; Social Networks; Trust, Epistemic

Further Readings

Baier, A. C. (1995). *Moral predjudices: Essays on ethics.* Cambridge, MA: Harvard University Press.

Baker, J. (1987). Trust and rationality. *Pacific Philosophical Quarterly, 68,* 1–13.

Castelfranchi, C., & Falcone, R. (2010). *Trust theory a socio-cognitive and computational model* (Wiley Series in Agent Technology). Chichester, England: Wiley.

Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity.* New York, NY: Free Press.

Gambetta, D. (Ed.). (1988). *Trust: Making and breaking cooperative relations.* New York, NY: Blackwell.

Hardin, R. (2002). *Trust and trustworthiness.* New York, NY: Russell Sage Foundation.

Jones, K. (1996). Trust as an affective attitude. *Ethics, 107,* 4–25.

Lagerspetz, O. (1998). *Trust: The tacit demand.* Dordrecht, Netherlands: Kluwer Academic.

O'Neill, O. (2002). *A question of trust: The BBC Reith lectures.* Cambridge, England: Cambridge University Press.

Tuomela, M. (2006). Rational social normative trust as rational genuine trust. In L. Siegal (Ed.), *Philosophy and ethics: New research* (pp. 1–56). Hauppauge, NY: Nova Science.

Truth, Philosophical Theories of

There is probably in philosophy no other notion that enjoys the privilege of being both so simple and so entangled. On the one hand, truth is, as René Descartes said, "so transcendentally clear that it is impossible to ignore it." On the other hand, as soon as we try to spell out the nature of the property or relation in which truth consists, we encounter difficulties, taking us to the highest reaches of metaphysics—those of the nature of knowledge, of the mind dependence or independence of reality, and of language and its relation to the world. This has led many thinkers to adopt Pontius Pilate's stance when he asked, shrugging his shoulders, "What is truth?" But Pilate was wrong. There is something important to say about truth.

In this entry, classical and more modern theories and notions of truth are critically reviewed. It is also pointed out that truth has a special normative role.

Classical Definitions of Truth

Can truth be defined? Most classical philosophers distinguish real definitions, which define the essence of a thing, from nominal definitions, which characterize adequately its concept. Aristotle's definition of truth seems to provide both: "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true"? (*Metaphysics*, Γ 7, 1011b 26–27). The tradition has taken it as a statement of what has come to be known as a *correspondence theory of truth,* according to which truth is a relation of correspondence between our judgments and reality.

Aristotle, however, does not say much about what this relation of correspondence might be, so it is tempting to interpret him rather as giving us only an explanation of the meaning of the word *true*. Similarly, when St. Thomas Aquinas says that truth is the *adaequatio* between *res* and *intellectus* (i.e., "reality" and "mind"), he does not tell us how to define the agreement in question.

For Immanuel Kant, truth as correspondence is only the "nominal" definition of truth, and he gives an argument against any attempt at giving a real definition of truth as correspondence: "I can only compare my judgments with an object by making a judgment about the object, so my judgment is correct only if it is confirmed by itself" (cited in Kunne, 2003, pp. 126–127). Although the argument is flawed (it conflates giving a definition of truth with giving a criterion of truth), it has had some success, and it has inspired Gottlob Frege's argument against the correspondence theory: If truth could be defined, a definition of it would have to say that *it is true* that such and such a property is ascribable to truth, hence would presuppose what is in question.

This objection also plagues many definitions of truth as correspondence to facts or states of affairs. If a proposition *P* is true if and only if it corresponds to the facts, what fact other than *the fact that P* itself can be a candidate for the corresponding entity? Thus, if it is true that *Rome is north of Naples,* it seems to be made true by the fact that Rome is north of Naples and also by the fact that Naples is south of Rome, but furthermore, it seems, also by the fact that Rome is to the north of the largest city within 20 miles of Ischia and such that London is in England. Clearly, since these descriptions have the same extension (i.e., refer to the same things: cities), they designate the same fact. So either we are bound to take *the fact that P* as trivially equivalent to *P*—in which case adding that the latter corresponds to the former is trivial—or we have to accept that each particular fact corresponds to potentially *all* the facts. This argument, known at the "slingshot," threatens all definitions of truth as correspondence to facts.

One natural reaction to the idea that there is nothing more in the notion of fact than that of a true proposition is to try to define truth as an epistemological notion. The *coherence theory of truth* says that a judgment is true if and only if it coheres with a system of other judgments. Coherence, however, can be defined in many ways—as a logical relation or as an explanatory one. Moreover, any addition of a belief to (or subtraction from) a coherent set of beliefs can destroy its coherence. To idealize by supposing that there is a maximally coherent set of all coherent sets quickly leads to versions of absolute idealism.

The second kind of epistemological theory is *verificationism*. It says that a judgment is true if and only if it can be warranted or justified, and so it identifies truth with our knowledge of truth. Most versions of verificationism grant that truth cannot be defined through actual verification: It is quite easy to imagine statements that are warranted but untrue. So most epistemic views of truth side for some version of idealized verification, according to which truth has to be in some sense ideally knowable. But at least some versions of this view lead to the "knowability paradox": Since all truths are knowable and since no one can know that a proposition is both true and not known to be true, it cannot be true that a proposition is both true and not known to be true, hence all truths are known, which is a *reductio* of verificationism.

Pragmatism can be also considered as a version of ideal verificationism. In its simple version, it says, with William James, that "truth is the expedient in our thinking" and that a true statement is one that is useful. Since this is easily refutable (many truths are useless, and many useful judgments are false), we are invited to idealize again and to consider truths in the long run, or at the end of inquiry. But such a state is hard to figure out.

Deflationism

Many philosophers have been tempted to conclude from the difficulties of formulating a satisfactory philosophical theory of truth that there is none to be had. Do we actually need to define truth? Our ordinary concept seems to consist of a set of trivialities: A true statement is true to the facts; truth is not the same thing as justification; truth is objective, is timeless, and has no degrees; and a statement "*P*" is true if and only if *P*. The latter platitude has been called the principle of "disquotation": It allows us, from *"P" is true,* to remove the quotes and the predicate *true* to get the equivalent assertion of *P*. What more is there when one says that it is true that snow is white than the assertion that snow is

white? On this view, *true* is just a logical word, like *and* or *not:* Truth has no essence and is not a deep metaphysical notion. *Deflationism* or minimalism is the doctrine according to which there is nothing more to say about truth than these trivialities. Alfred Tarski gave, in this spirit, a semantic theory for formal languages in which truth is defined, for a given language, in a metalanguage in which one can derive equivalences of the form *P* is true if and only if *P.* Tarski proved that truth cannot be defined *within* a language but has to be defined in another language, and there is no limit to the hierarchy of languages.

Deflationism, however, is hardly tenable. First, if it says that truth is a predicate of sentences, it is implausible, since truth is a predicate of propositions, or of thought contents. We can certainly say that a sentence like "$E = mc^2$" is true even when we do not understand what it means, but we can hardly say that we have asserted a truth if we do not know what truth it is. Second, truth is more than a mere device of disquotation. It is also a norm of assertion (if one asserts that *P,* then *P* is supposed to be true) and a norm of belief (one ought to believe only what is true). If we disregard this normative dimension, we obviously miss something of the point of the concept of truth. Third, the deflationist concept of truth cannot capture the objectivity that is involved in truth. If truth is reduced to assertion, any kind of assertion will aim at truth and any kind of discourse to which the notion of truth can be applied will be equally susceptible to being true, independent of the reasons that one has to assert. Truth seems to apply to a number of distinct domains: There are mathematical truths, physical truths, moral truths, legal truths, historical truths, sociological truths, possibly metaphysical and religious truths, but also fictional truths, poetical truths, and comic truths, and so on. If truth is but a device of assertion, shall we say that it applies equally to all these domains? But certainly we do not want to say that statements in fictional narratives, for instance, are just as true as mathematical ones or that truth aptness is the same for physics or for ordinary objects, such as tables and cars, and for ethics or literature. The deflationist theory seems too hospitable. If we are ready to accept that there are ethical truths, it is not clear that we want to say that they are true in the same sense as the truths of physics. This minimalist conception of truth actually deprives the concept of truth of any bite. More than that, if truth is just the expression of one's opinion, when one asserts that *P,* then what prevents us from saying that there are as many truths as opinions, hence falling into complete relativism?

Realism Versus Relativism

We face a dilemma. On the one hand, we want to say that truth is a concept or a property that is robust enough to include a set of features associated with a realistic position: True judgments must correspond to an independent reality, which in some sense causes them rather than the reverse, on which there can be intersubjective agreement, and they must be stable and noncontextual. On the other hand, we want to allow that truth can be less objective in some domains, without losing its central properties. If we take the first horn, we face the problem of having to define the appropriate notions of correspondence, fact, and objectivity, which is not an easy matter; and we risk a kind of truth chauvinism, which will restrict the application of truth only to very few domains (are we even sure that the realist concept applies fully in physics?). If we choose the second horn, we risk ending up with a too welcoming, but excessively shallow, notion of truth.

One way out of this dilemma is to defend a kind of *functionalism* about truth, in analogy with the corresponding doctrine in the philosophy of mind. Just as the mental state of pain can be defined by its causal role and realized in various ways depending on the organisms that instantiate it (e.g., in different neuronal configurations in mammals, reptiles, or cephalopods), we can say that truth is defined by its role as characterized by its formal properties and the various truisms associated with it, although these properties are realized differently in various domains. Thus, truth might not be correspondence in certain domains, such as ethics or literature; or it could be ideal warrant in some other domains, such as mathematics, or a different kind of property in the domain of fiction. We could thus allow a form of truth pluralism, together with an acceptance of a core of truth properties applicable in different ways to distinct domains. But the common functional core of truth properties must not be too formal, for we risk being led back to the deflationist view. We need also to be able to admit that truth does *not*

apply in certain domains, or that if it applies, it is sufficiently robust. Take the case of comic truth. You say that Buster Keaton is funny, and I say he isn't. Comic truth is highly relative. So is it truth at all? Suppose, however, that, along with a kind of classicism in aesthetics, we accept that there are some minimal canons of the comic. We could thus accept what David Hume calls a common standard of taste, and some consensus on what is funny, although we would deny that we can reach agreement as in, say, history. We would have to say that the notion of truth is very etiolated in this domain, or perhaps that it does not apply at all. The core properties of truth must involve at least some claims about the possibility for truth to determine some kind of *knowledge* of independent facts, and not merely rational beliefs, in a realist sense of this notion: This entails that some truths might not be known. Such a concept probably excludes the comic or the fictional from the realm of truth aptness.

For the same reasons, there cannot be relative truths. Indeed many, if not most, truths are contextual, in the sense that they have to be evaluated relative to given situations, times, places, speakers, and so on. If I say that I am hungry, the truth of my statement is relative to me, to the present moment, and so on. But once these parameters are fixed, the statement is perfectly evaluable for truth. Cultural relativism is the view that our judgments are relative to various frameworks—to a language, to a conceptual scheme, to a community, or to various historical circumstances. But can the idea that the schemes or frameworks be incommensurable be made to work? According to a well-known argument by Donald Davidson, this is impossible if we use the concept of truth in the ordinary sense.

Can there be nevertheless some faultless disagreements in some domains, such as in matters of taste or in ethics? If I say that that murder is wrong and if you say that it is not, can we say that my statement is true for me and yours true for you, hence that we do not contradict each other? The situation is better described as one in which I *believe* that murder is wrong and you *believe* that it is not, which is an ordinary disagreement. That certain statements can be assessed only relative to a circumstance of assessment is undeniable, but does it follow that there can be a concept of "True for x"? This is dubious, since it would entail that a statement could be both true and correct for x but not true and correct for y. But the norm of correctness for truth would disappear if it were made relative to a perspective.

The Value of Truth

Truth is a norm for our assertions and our beliefs: We aim at having true beliefs, and if we discover that one of our beliefs is false, we have to reject it. Does it follow that we ought to believe all truths, including those that are of no interest for us? If we believe that P, and if Q logically follows from P, but Q is absurd, are we under the obligation to believe Q? Normative requirements, like the norm of truth or the principles of logic are general; they do not entail that we should believe all truths and infer all consequences without attending to the reasons for our beliefs. The same holds about the value that is attached to truth. Truth is the aim of scientific inquiry, and most of the transfer of information within a society aims at propagating true beliefs.

Democracy could thus not exist, and not only if the basic rights of freedom of opinion and of speech were not granted, and thereby the capacity to spread information and to evaluate it were blocked. Nevertheless, this social and political primacy of truth has been contested on at least three fronts. Thinkers like John Rawls or Jürgen Habermas have argued that a politically liberal society cannot rest on the notion of truth, since it would entail that some political arrangements are true whereas others are not, and they have argued that weaker notions such as rational consensus are better. Sceptics or relativists about truth, from Friedrich Nietzsche to Michel Foucault, and more recently Richard Rorty, have argued that invoking the notion of truth in politics hides an attempt to confiscate power in the name of a spooky ideal, and they have proposed to replace truth by notions such as solidarity or wisdom. And a number of writers in science studies, too, have argued that truth and objectivity are but the masks under which political power hides itself.

It is not clear, however, that attempts to replace truth by other, more politically useful notions can actually dispense with the concept of truth. If democracy is not to fall into what John Stuart Mill called the tyranny of opinion, one needs an objective concept of truth, whereby it is possible to give reasons in favor of the truth of certain views. The fact that history and the contemporary observation

of the media culture shows that people *claim* their views to be true but aim at manipulating opinion does nothing to show that truth itself is a fiction in the service of political power. On the contrary, without the capacity to say that "two plus two equals four," we might end up in a world not so different from the one of Orwell's *Nineteen Eighty-Four.*

Pascal Engel

See also Empiricism; Epistemology; Falsifiability; Idealism; Induction and Confirmation; Metaphysics and Science; Objectivity; Pragmatism; Relativisms and Their Ontologies; Science and Ideology; Verificationism; Virtue Epistemology

Further Readings

Blackburn, S. (2005). *Truth: A guide.* London, England: Penguin.

Crivelli, P. (2004). *Aristotle on truth.* Cambridge, England: Cambridge University Press.

Davidson, D. (1984). *Inquiries into truth and interpretation.* Oxford, England: Clarendon Press.

Dummett, M. (1959). Truth. *Proceedings of the Aristotelian Society, 59,* 141–162.

Engel, P. (2002). *Truth.* Durham, England: Acumen.

Frege, G. (1956). The thought. *Mind, 65,* 289–311.

Goldman, A. (1999). *Knowledge in a social world.* Oxford, England: Oxford University Press.

Horwich, P. (1998). *Truth* (2nd ed.). Oxford, England: Oxford University Press.

Künne, W. (2003). *Conceptions of truth.* Oxford, England: Oxford University Press.

Lowe, E. J., & Rami, A. (2009). *Truth and truth-making.* Stocksfield, England: Acumen.

Lynch, M. P. (2009). *Truth as one and many.* Oxford, England: Oxford University Press.

Richard, M. (2008). *When truth gives out.* Oxford, England: Oxford University Press.

Tarski, A. (1983). *Logic, semantics, metamathematics.* Indianapolis, IL: Hackett.

Williams, B. (2002). *Truth and truthfulness.* Princeton, NJ: Princeton University Press.

Wright, C. (1993). *Truth and objectivity.* Oxford, England: Oxford University Press.

UNCONSCIOUS

In ordinary language, "unconscious" is often used to characterize the condition of having no awareness, as when one is in a coma, or in a deep, dreamless sleep. In contrast, psychologists use the term *unconscious* to refer to a characteristic of certain mental states. A mental state is unconscious if and only if (a) one is in that state and (b) one is not aware that one is in that state. There is widespread agreement among social and behavioral scientists that a great deal of our mental life is unconscious and that any adequate explanation of human behavior must include the unconscious within its purview. However, there are fundamental disagreements about the nature of unconscious mental states and their relation to consciousness.

Historical Background

Although notions of unconscious mental processes have a pedigree extending back to the ancient world, the unconscious was not seriously taken up by science until the 19th century. To understand what precipitated this development, it is necessary to consider its historical background.

The 17th-century French philosopher René Descartes argued that consciousness is a *defining characteristic* of the mental; nothing can be mental unless one is conscious of it. The notion that mentality is coextensive with consciousness was immensely influential and was widely accepted by philosophers and scientists during the ensuing three centuries. This conception of the mind came under increasing pressure during the latter half of the 19th century, when the newly minted sciences of the mind (neurology, psychology, and psychiatry) began to accumulate observations that were difficult to reconcile with the Cartesian paradigm. For example, experiments using hypnotism demonstrated that behavior can be caused by mental states of which one is unaware. Subjects can be given a post-hypnotic suggestion to perform some action in response to a trigger after they have emerged from the hypnotic state, *without being aware of why they are doing this*. Observations of this kind—as well as of phenomena such as anosognosia (a neurological disorder characterized by unawareness of an obvious disability)—and the discovery of the role of nonconscious inferences in visual perception, all militated against the view that the human mind is transparent to itself.

Researchers struggled to find ways to reconcile these phenomena with the Cartesian conception of the mind. One strategy, sometimes called *dissociationism,* was to deny that so-called unconscious mental states were really unconscious. Advocates of this approach were influenced by studies of what is nowadays called "multiple-personality disorder" (or "dissociative identity disorder"), a form of mental illness in which two or more distinct, alternating personalities (or "selves") inhabit the mind of the sufferer. Dissociationists proposed that one or more dissociated secondary consciousnesses can exist alongside a person's primary consciousness. According to the dissociationists, ostensibly unconscious mental states are states of a split-off secondary consciousness to which one's primary

consciousness does not have access. A different explanatory strategy, sometimes called *dispositionalism,* was to deny that unconscious states are really mental. To understand it, one needs to understand what dispositions are. Dispositions are latent tendencies: A thing has the disposition to *F* if it tends to *F* under certain circumstances. The property of fragility is a good example of a disposition. To say that an object is fragile is to say that it has the disposition to shatter easily. One of the important features of dispositions is that they can fail to give rise to the states that they are dispositions for (a fragile wine glass may never shatter, because it is never exposed to conditions under which the disposition to shatter would be realized). The dispositionalists' view of unconscious mental states was that specific states of the central nervous system dispose one to have specific conscious mental states. The idea is that for any mental state *M,* there is some neural state *N* that is the disposition for *M* and without which *M* could not occur. However, just like the wine glass, it is possible for a brain to be in *N* without being in *M*. Dispositionalists held that being in such neural states can affect behavior in ways that are similar to the effects of the corresponding mental states. Because of this, it is tempting—although erroneous—to think of them as unconscious *mental* states.

Both dissociationists and dispositionalists struggled to make sense of recalcitrant facts while remaining loyal to the Cartesian presumption that all mental states are conscious. Sigmund Freud's conception of the unconscious is philosophically significant precisely because he jettisoned the Cartesian presumption in favor of the view that cognition is essentially unconscious.

Freud

For many people, the notion of the unconscious is inextricably linked to the work of Sigmund Freud (1856–1939). Freud was one of the first students of human nature to provide a sophisticated theory of the unconscious aspects of mental life. His work has had an immense impact on the social and behavioral sciences, although the extent of this influence is often unrecognized and unacknowledged. Trained as a clinical neurologist, Freud was familiar with discussions about the unconscious that were unfolding in the scientific literature during the closing decades of the 19th century. Although early in his career he embraced the view that all mental states are conscious, he abandoned this position in 1895 in favor of the view that our psychological life is mainly unconscious.

Freud's theory of the unconscious resists easy summary, in part because of its complexity and in part because he reconfigured it several times during his lengthy career. To explain it, it is helpful to begin with his account of consciousness. Freud adhered to what is nowadays known as a *restrictivist* theory of consciousness—that is, he believed that all conscious mental states have a sensory or "qualitative" character. He argued that even very abstract thoughts have a qualitative component, because they are disclosed to consciousness in the form of inner speech. When you entertain a conscious thought—for instance, the thought that there is a highest prime number—what goes on in your mind resembles the experience of *hearing your own utterance* of the sentence "There is a highest prime number." Of course, you do not literally hear this sentence, but you have an experience that is something like hearing it. Freud explained this by hypothesizing that cognitive states are unconscious and become conscious only by activating what he called "motor speech representations" (the neurological basis for speech) that express their content. These outgoing impulses are too weak to produce vocalizations, but they generate feedback to the brain that gives rise to the quasi-auditory sensations of conscious thought. So cognition is an unconscious process. As Freud often put it, deliberately using a Kantian idiom, the mental is unconscious *an sich* ("in itself"). *All* of our thoughts occur outside of awareness. What we call conscious cognitions are really just conscious *representations* of unconscious cognitive states.

If consciousness plays no essential role in cognition, then it is possible that there are unconscious mental states that never become conscious. Freud argued that this is the case and held that it can occur in two ways. Many of our cognitive states do not become conscious because their becoming conscious would make no contribution to our lives. These states are unconscious in consequence of our cognitive architecture. Other states are actively barred from entering consciousness—they are, in Freud's terminology, "censored" or "repressed." Repressed thoughts are thoughts that are not represented (or, alternatively,

misrepresented) in inner speech. Freud believed that repression is emotionally driven: Thoughts are excluded from consciousness in virtue of their capacity to produce severe psychological conflict.

It is often said that Freud claimed that we repress emotions and sexual impulses. In fact, he specifically repudiated the idea that emotions and impulses can as such be repressed. In Freud's theory, it is only thoughts—that is, propositionally structured cognitive representations—that are capable of being repressed. We can speak loosely of repressed affective and appetitive states, but we must bear in mind that it is the thoughts associated with those states that can be repressed, rather than the states themselves. Freud thought that the repression of such thoughts has a significant impact on behavior, because thoughts excluded from consciousness cannot function as reasons for one's actions. So repressing the cognitive component of an emotion or desire effectively blocks its direct expression as a motive for action (although not its indirect, nonrational expression).

During the final decade of his life, Freud was especially concerned with the interplay between individual psychology and cultural life. He argued that participating in the cooperative enterprise of civilization requires us to repress (in the loose sense described above) certain desires that are hostile to the social order. Culture demands repression and enforces it through child-rearing practices that shape psychological development in socially congenial ways. Conversely, repression also feeds culture by diverting our primitive, antisocial desires into socially acceptable, symbolic outlets (sublimations). Freud held that cultural practices and institutions, such as art, religion, and even science, are ultimately motivated by, and depend on, repressed psychological forces. The explanatory relation between repression and culture is reciprocal: Repression cannot be properly understood unless one takes culture into account, and culture cannot be properly understood unless one takes repression into account.

Post-Freudian Developments

Although psychoanalytic theory had an immense impact on conceptions of human nature, it came to be regarded as a scientific failure. Psychoanalytic claims about how the mind works were, although intuitively appealing, empirically unsupported and experimentally intractable. As the 20th century progressed, psychology became an increasingly rigorous scientific discipline and had little use for the speculative claims of Freudian theory.

For much of the century, psychology was dominated by versions of behaviorism. Although behaviorism provided a powerful theory of how learning shapes behavior, it was also wedded to a scientific ideology that eschewed any talk about inner mental processes. The behaviorist monopoly was eventually broken during the 1960s by what became known as the "cognitive revolution." Inspired by the rise of computer science, cognitive scientists began to focus on the subpersonal information-processing routines that underpin human behavior: information-processing routines that are structurally unconscious. Because of the Freudian connotations of the term *unconscious,* psychologists at first tended to refer to these processes as "automatic" or "nonconscious." However, in 1987, the psychologist John Kihlstrom, also a contributor to this encyclopedia, wrote an influential paper in which he argued that much of the research on cognitive science is concerned with what he called the "cognitive unconscious," borrowing a term coined by Jean Piaget. Although Kihlstrom described the Freudian unconscious as a version of the cognitive unconscious, psychologists have tended to use the term *cognitive unconscious* to differentiate their perspective from the Freudian one. Research into the cognitive unconscious has given rise to an immense scientific literature and a plethora of innovative experimental procedures. Although the existence of unconscious mental processing is no longer controversial in psychology, there is still debate about how sophisticated unconscious cognition can be—that is, whether the unconscious is "smart" or "dumb." Investigations into the subliminal perception stimuli, memory, learning (especially language acquisition), emotion, and motivation have tended to support the hypothesis that quite complex cognitive processes can, and often do, occur outside of awareness.

David Livingstone Smith

See also Consciousness; Ego; Personal Identity and Trauma; Philosophical Psychology, History of; Psychoanalysis, Philosophical Issues in; Schizophrenia: Psychoanalytic, Phenomenological,

and Contemporary Philosophical Approaches; Therapy, Psychological and Philosophical Issues; Unconscious Social Behavior

Further Readings

Ellenberger, H. F. (1981). *The discovery of the unconscious: The history and evolution of dynamic psychiatry.* New York, NY: Basic Books.

Freud, S. (1957). The unconscious. In *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 14, pp. 159–215). London, England: Hogarth Press/Institute of Psychoanalysis. (Original work published 1915)

Hassin, R. R., Uleman, J. S., & Bargh, J. A. (2005). *The new unconscious.* New York, NY: Oxford University Press.

Kihlstrom, J. F. (1987). The cognitive unconscious. *Science, 237,* 1445–1452.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious.* Oxford, England: Oxford University Press.

Reber, A. S., & Allen, R. (1999). Unconscious intelligence. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (pp. 314–324). Oxford, England: Blackwell.

Smith, D. L. (1999). *Freud's philosophy of the unconscious.* New York, NY: Springer.

Wilson, T. D. (2004). *Strangers to ourselves: Discovering the adaptive unconscious.* Cambridge, MA: Harvard University Press.

Unconscious Social Behavior

Psychological explanations attribute social behavior to the individual's mental states—the thoughts, feelings, and desires that cause the individual to act in a particular way toward others. While the cognitive revolution in social psychology appeared to emphasize the role of conscious cognition in social interaction, a recent trend has been to underscore the role of *unconscious* cognitive processes automatically evoked by stimulus inputs and executed outside phenomenal awareness and voluntary control. However, claims that social behavior is dominated by unconscious processes, such that "free will" plays little or no role in human experience, thought, and action, go beyond the presently available evidence.

The earliest psychological theories of social behavior, put forward in the 1920s and 1930s, were couched in a version of stimulus–response behaviorism that made no reference to consciousness. This attitude carried over into the "golden age" of experimental social psychology, in the 1950s and 1960s, which emphasized social influence and the power of the situation to influence the individual's experience, thought, and action. However, the cognitive revolution in experimental psychology gave rise to an alternative cognitive perspective, which emphasized people's *perception* of the situation, their goals, their expectations concerning the consequences of their actions, and so on. Although it was rarely stated quite so clearly, the implication of the cognitive perspective in social psychology—like the cognitive perspective in psychology generally—was that the percepts, memories, and thoughts that mediated social behavior were consciously accessible to the actor, guiding his conscious choices and actions.

Beginning in the 1970s, however, cognitive psychology reawakened an interest in unconscious mental life with the distinction between *automatic* and *controlled* processing. Controlled processing is conscious and deliberate; it consumes cognitive resources and involves serial processing. Automatic processes, by contrast, are inevitably evoked by the appearance of particular environmental stimuli; once evoked, they are incorrigibly executed, in a "ballistic" fashion; they consume few or no cognitive resources; and they do not interfere with each other or with controlled processes, thus permitting some degree of parallel processing. Automatic processes are reflex-like in some respects, but they are not necessarily innate: In principle, any process, no matter how complex, can be automatized if it is practiced diligently enough. Whether they are innate or acquired, automatic processes are unconscious in the strict sense of the term; they operate outside conscious awareness and are independent of conscious control.

The automatic/controlled distinction was quickly imported into social psychology, with a number of prominent investigators arguing that much of social behavior occurs automatically in response to certain cues, without mediation by conscious, deliberate thought. Within cognitive psychology, there is a general consensus that every task has both automatic and controlled components, and considerable effort has been devoted to measuring their differential

contributions to performance. In social psychology, however, a view has developed that social cognition and behavior are overwhelmingly governed by automatic processes—with one theorist invoking classic "Ivory Soap" advertisements to assert that behavior is 99.44% automatic and another asserting that that free will is so severely compromised by automaticity that, *pace* Descartes, we are automatons after all.

Automaticity has been dubbed "the new unconscious"—the "old" unconscious being the "monsters from the Id" (a phrase from the 1956 science fiction film *Forbidden Planet*), envisioned by Sigmund Freud and other proponents of classical psychoanalysis. But the basic idea was anticipated by William McDougall's "hormic psychology," which argued that social behavior was motivated by a set of basic instincts that operated unconsciously, as well as by the "behavior viewpoint" of Floyd Allport, that social behavior reflected unconditioned or conditioned responses to social stimuli. The embrace of automaticity does not exactly revive Skinnerian behaviorism, however, because the new theorists adopt the central dogma of the cognitive revolution—that cognitive, emotional, and motivational states and processes intervene between the environmental stimulus and the organismal response. But when the intervening states and processes are automatically evoked by environmental stimuli, the embrace of automaticity looks more and more like behaviorism with a cognitive face.

In fact, nothing in the literature justifies the assertion that automatic processes dominate social cognition and behavior. Most of the published research constitutes demonstration experiments that merely show that automaticity plays *some* role in social interaction. But many of these experiments involve a very loose operationalization of automaticity, relying on fewer than the four canonical features listed above. The few comparative experiments published to date reveal much more of a balance between the automatic and controlled components of processing—except in special circumstances, such as very narrow response windows, where controlled processing simply cannot come into play. Nothing in the literature supports the idea that social behavior is wholly, or even largely, driven by automatic processes—much less, as some have concluded, that conscious will is an illusion and introspection only gets in the way of adaptive behavior.

Acceptance of the concept of automaticity has helped legitimize the concept of unconscious mental life, but it does not exhaust the possible unconscious determinants of social behavior. In the conventional view, automatic processes operate on conscious mental contents—percepts, memories, thoughts, and the like—to generate other conscious mental contents. We are aware of *what* we think, even if we are not aware of *why* or *how* we think it. Beginning with the study of implicit memory in amnesic patients, however, it has become clear that mental states—percepts, memories, the knowledge acquired through learning—can influence ongoing experience, thought, and action in the absence of phenomenal awareness. Indeed, the "Implicit Association Test" has been promoted as a means of assessing unconscious attitudes and beliefs that can result in prejudice and aggression directed toward social out-groups. Although the idea of unconscious beliefs, attitudes, and goals remains controversial, widespread acceptance of implicit memory, perception, and learning in the cognitive domain implies that the notion of implicit emotion and motivation should not be dismissed out of hand. Because the lack of conscious awareness precludes conscious control almost by definition, any effects of unconscious thoughts, feelings, and desires must be mediated by automatic processes.

John F. Kihlstrom

See also Behaviorism in Psychological Explanation; Consciousness; Determinism; Free Will in the Social Sciences; Implicit Bias and Social Cognition; Unconscious

Further Readings

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin, 126*(6), 925–945.

Kihlstrom, J. F. (2008a). The automaticity juggernaut. In J. Baer, J. C. Kaufman, & R. F. Baumeister (Eds.), *Psychology and free will* (pp. 155–180). New York, NY: Oxford University Press.

Kihlstrom, J. F. (2008b). The psychological unconscious. In O. John, R. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 583–602). New York, NY: Guilford Press.

McDougall, W. (1950). The Hormic psychology. In *An introduction to social psychology* (Suppl. chap. 7). London, England: Metheun.

Wegner, D. M. (2002). *The illusion of conscious will.* Cambridge: MIT Press.
Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious.* Cambridge, MA: Harvard University Press.

UTOPIANISM

Utopianism is a term used across the humanities and the social sciences to refer to dreams of a better life or to orientations that transcend what is currently possible or realistic. The term developed in the context of European accounts of utopia.

Meaning

Utopia means "no place" (from *topos,* Greek for "place" and the privative prefix *un-*), especially an ideal place that is no place—that is, no actualized or real place—although it sometimes extends to spaces in which we can be different. Some, such as Leszek Kołakowski, want to define utopia narrowly as referring only to beliefs that a definitive and unsurpassable condition is attainable that can be arrived at by human efforts. Others associate utopia with dreams of a better life quite generally, and so with the human capacity to envisage alternative possibilities. On this broader view, utopias allow thinking beyond the present; they also manifest human desire and offer critiques of existing states of affairs.

Types

Utopias are of many different types. They include literary texts; accounts of ideal societies, especially societies based on social harmony; social designs of allegedly better arrangements for practical life; social and economic outlooks; works of political and social theory; and intentional communities. There are technoscientific utopias, green utopias, gender utopias, geographic utopias, electronic and digital utopias, and so forth. Utopias may be realistic or imaginary; eutopic, heterotopian, or dystopian; on the left or the right; and authoritarian or libertarian. They may be based on an allegedly perfect human nature, or they may be about how imperfect human nature can be changed or managed. There may also be utopias of space and time. Distinctions are also possible with utopias made by human beings and those not so made. Utopia also has a history, and the manifestations of utopianism vary over time; they include some forms of millenarianism as well as ideal lands such as Arcadia and Cockaigne.

The Study of Utopianism

The study of utopianism, an area that has developed as a serious area of inquiry in recent years, illustrates the need for *a closer relationship between philosophy and the social sciences.* The standard charges that utopias are unrealistic, impractical, and potentially dangerous are overstated and only partially supported by detailed studies of intentional communities such as religious orders, communes, and kibbutzim. Some utopias are realistic and/or have practical elements. Clearly, utopias that are elements of political movements can be dangerous, as attempts to realize utopian designs in practice may be in any area. Nonetheless, the deeper significance of utopia and utopianism is not restricted to the sociopolitical but extends to the metaphysical and natural-scientific as well. Older work on utopia tends to neglect both and confuses utopianism in general with European utopianism, with its focus on nonexistent ideal societies and the pursuit of social harmony and happiness. While it is now widely recognized that utopian thinking may be necessary in a range of circumstances for human beings, there is little contemporary philosophical or natural-scientific work on why this should be so.

Social Theory

This situation partly reflects the tendency for different disciplines to understand utopia and utopianism differently. Sociologists, following Karl Mannheim, who famously contrasted ideology and utopia, associate utopia with the emergence of modern society in Western Europe and with models of social harmony. They sometimes forget that utopianism has flourished in many parts of the world and before the modern period, especially in Asia. In cultural sociology, utopia is used to describe the nonplaces of the consumerism generated by capitalism. In politics, utopianism is associated with the problems of perfectionist politics and communism.

Following the collapse of communism and the attacks by Isaiah Berlin and Leszek Kołakowski implying that utopia is based on ethical monism and leads to repression and violence, utopianism has fallen out of fashion, at least at the level of a political strategy.

On the other hand, although many writers refer to a decline or even an end of utopia, neoliberalism

has flourished, despite its utopian elements, even though those advocating such utopianism have often denounced utopia as the left-wing illusion that it is possible to achieve goals such as equality and social justice. There has also been a retreat from the redemptive utopianism associated with German Jewish thinkers such as Walter Benjamin, Herbert Marcuse, and Theodor Adorno, a development partly explained, as the American Marxist Frederic Jameson notes, by changes in capitalism. In law, utopianism tends to be practical and about designing better institutions, organizations, and regulative regimes. The distinguished Brazilian legal theorist Roberto M. Unger advocates a utopianism of this sort. Given the contemporary need to devise new global institutions, utopianism as a form of practical political philosophy and a technique of invention probably has a future more than most political commentators suggest.

Philosophy

In philosophy, utopianism is associated with technical issues in metaphysics and logic, including theories of possibility and possible worlds. Philosophers are more inclined than social scientists to distinguish the *ideal,* or what is normatively maximal, from what is *without a place*—and both from what is *only imagined*. Utopian philosophy and the philosophy of utopianism, which is often a subset of it, has been relatively neglected by scholars of utopianism at the technical level. The most important utopian philosopher is the German Jewish philosopher Ernst Bloch (1875–1977). Bloch reinterprets utopia as a feature of reality itself and as having a place in the now of the moment. In his masterpiece *The Principle of Hope* (three volumes), he argues that utopian surplus can be found throughout human consciousness, in daydreams and in cultural materials from all over the world, and also in nature. Bloch's work has been noted by scholars of utopia, but his philosophical contributions, including his utopian metaphysics, have been largely overlooked. French discussions of philosophy and utopianism (by Louis Marin and Michele Le Doeuff), on the other hand, have been taken up to some extent.

In the longer term, modern European utopianism is likely to be eclipsed by global utopianism, which takes the philosophy of possibility and nonexistence seriously. Students of utopia have too often ignored religious utopianism; even in religious studies utopia has been understudied, even though religious utopias are found worldwide and are often strikingly successful over many centuries. This is now changing with the development of postsecular thought, and many contemporary scholars interested in utopia are attempting to rethink the relationship between the secular and the religious. Once again, serious attention is only now being paid to the ontological implications of the tendency of the human imagination to project counterfactual states of affairs and to the biology that makes it possible for them to do so.

Wayne Hudson

See also Enlightenment, Critique of; Marxism and Social/Historical Explanation; Philosophy of History

Further Readings

Adorno, T. (1973). *Negative dialectics* (E. B. Ashton, Trans.). London, England: Routledge & Kegan Paul.

Benjamin, W. (1999). *The Arcades Project* (H. Eiland & K. McLaughlin, Trans.). Cambridge, MA: Belknap Press.

Berlin, I. (1991). *The crooked timber of humanity: Chapters in the history of ideas*. London, England: Fontana.

Bloch, E. (1954–1995). *The principle of hope* (3 vols., N. Plaice, S. Plaice, & P. Knight, Trans.). Cambridge: MIT Press.

Goodwin, B. (Ed.). (2001). *The philosophy of utopia*. London, England: Frank Cass.

Hudson, W. (1982). *The Marxist philosophy of Ernst Bloch*. New York, NY: St. Martin's.

Hudson, W. (2003). *The reform of utopia*. London, England: Ashgate.

Jameson, F. (1988). *Utopia post utopia*. Boston, MA: Institute of Contemporary Art.

Kołakowski, L. (1990). The death of utopia reconsidered. In *Modernity on endless trial* (pp. 131–145). Chicago, IL: University of Chicago Press.

Le Doeuff, M. (1989). *The philosophical imaginary* (C. Gordon, Trans., chap. 4). Stanford, CA: Stanford University Press.

Mannheim, K. (1979). *Ideology and utopia: An introduction to the sociology of knowledge*. London, England: Routledge & Kegan Paul.

Marcuse, H. (1970). *Five lectures*. London, England: Allen Lane.

Marin, L. (1984). *Utopics: Spatial play* (R. Vollrath, Trans.). Atlantic Highlands, NJ: Humanities Press.

Sargent, L. (2011). *Utopianism: A very short introduction*. Oxford, England: Oxford University Press.

Unger, R. M. (1999). *Politics: The central texts* (Z. Cui, Ed.). London, England: Verso.

Value Neutrality in Science

The claim of the value neutrality of science is that *scientific* knowledge neither supports nor undermines any value judgment and that accepting a scientific theory in accordance with proper epistemic criteria is compatible with holding any viable value outlook. The claim is consistent with science being "value relevant": Scientific knowledge may inform the means to bring about goals that are judged valuable, and priorities of research may be set by the goal of producing such knowledge. Since Max Weber's influential writings in the early 20th century, defending the sound scientific status of the social sciences has often included arguments that the social sciences are, or ought to be, value neutral. This entry will first introduce the philosophical background of the value neutrality of science; then, after four criticisms of its applicability to the social sciences are sketched, its viability as an ideal will be briefly entertained.

Philosophical Background

The claim that science is value neutral follows from three proposals: first, the alleged dichotomy of fact and value, the proposal that factual statements and value judgments are distinct and separable, having no relations of logical entailment with one another; second, that science deals only with matters of fact; and, third, that scientific knowledge could support or undermine value judgments only if the latter were logically entailed by the former. The dichotomy of fact and value often is grounded in David Hume's famous argument that "ought" cannot be entailed by "is." It has been reinforced by the common views that value judgments express subjective preferences or emotional responses and (in the natural sciences) that scientific theories represent the underlying structures of phenomena, the processes and interactions of their components, and the laws that govern them, dissociated from their connection with human experiences, lives, values, and social/cultural forms—so that the language of scientific theory is "mathematical" or "technical" and so does not contain the categories needed to make value judgments. While attempts have been made—for example, in economics and behaviorist psychology—to deploy only "technical" categories, in the social sciences generally, this has not been a serious option.

Many criticisms made of the value neutrality of the social sciences begin with questioning the third proposal above. They do not question that scientific knowledge can have no logical entailments in the realm of values. Rather, they maintain that scientific knowledge gained in the social sciences can sometimes support or undermine value judgments in virtue of the relations that exist between factual statements and value judgments in certain contexts, which make it unintelligible to accept the factual statement and to deny a specified value judgment—unless explanation that makes sense of denying the value judgment is provided. In these contexts, the value judgments are not logically entailed by the relevant factual statements (or accepted theoretical explanations) but follow from them, other things being equal, thereby providing support for—that is, a reason to make—the value judgments. The first

three criticisms sketched below are of this kind; in them, the source of departure from neutrality is provided by these nonentailment relations. The fourth criticism locates the source differently, when—unlike most versions of the first three—it challenges part of the first proposal by questioning whether the criteria used in evaluating scientific knowledge are free from compromises with particular ethical/social values.

Criticisms

Social-Scientific Understanding "Secretes" Value Commitments

The adoption of a *theoretical framework* (Charles Taylor's term) is motivated by a conception of human wants, needs, and purposes, or reinforced by holding a particular value outlook that influences the kinds of categories used in efforts to gain understanding of phenomena and contributes to selecting the phenomena for which understanding is to be sought in inquiry. These categories are not only limited to "technical" ones but also include terms (sometimes called "thick ethical terms") that (unlike "thin ethical terms," e.g., *good* and *bad*) may be used simultaneously in descriptive and evaluative modes—so that value judgments follow, other things being equal, from empirically confirmed claims. Consider, for example, the following statement: "The drug companies are being driven more by financial ambition and marketing considerations than by scientific and public health objectives, and that is the root of their current problems." Clearly, this claim is subject to empirical test. However, it would be unintelligible, other things being equal, to accept that it is well confirmed and not to make the value judgment that the drug companies are acting badly—*driven more by financial ambition* is one of the thick ethical terms deployed in making the claim. This is an example of how social-scientific understanding *secretes* (Taylor's term) value commitments. The empirically confirmed claim does not logically entail the value judgment. There is no logical contradiction involved in accepting the claim and denying the appraisal. Other things might not be equal! But it is not intelligible to accept the claim and deny the value judgment, unless one provides evidence that other things are not equal (e.g., by providing evidence that, in current circumstances, responding to marketing considerations is the only way to get funding to support public health programs).

"Explanatory Critique" in the Social Sciences

Research in the social sciences may lead to the confirmation or disconfirmation of a proposition such as "Democracy is not viable outside the institutions of capital and the market." It may also lead to the confirmation of statements about who believes the proposition, about how widely the belief is shared, and, if it is widely believed, about the causes of its being so. This makes possible what Roy Bhaskar calls "explanatory critique." If research were to disconfirm the proposition, and also to confirm that it is widely believed (though false) and that its being widely believed is partly explained in terms of mechanisms integral to prevailing social structures, then—other things being equal—a negative value judgment of the social structures follows, and also—other things being equal—a positive value judgment of projects aiming to replace them with alternative structures. Again, other things might not be equal—for example, evidence might be available that the alternative structures would probably be the source of greater harm. Explanatory critique, according to Bhaskar, reflects an "emancipatory impulse" in the social sciences; the social sciences have the capacity to uncover "false consciousness" and its sources, and from such results follows, other things being equal, positive appraisal of projects aiming for emancipation from the grip of ideology. Unlike the first criticism, explanatory critique does not depend on any role played by "thick ethical terms."

Social Science Results May Undermine Presuppositions of Value Outlooks

Rejecting the view that value judgments express subjective preferences or emotional responses, some philosophers maintain that holding values presupposes—or follows, other things being equal, from—claims that are open to empirical investigation. Hugh Lacey illustrates this when he maintains that holding values of technological progress—for example, the value of technoscientific innovations introduced for the sake of economic growth—presupposes claims such as "Technoscientific innovation provides benefits that contribute toward the well-being of human beings generally"; "There are technoscientific solutions to most human problems, including those occasioned by technoscientific innovations themselves"; and "There are no serious alternative proposals available today to the pursuit of economic growth

based on technoscientific innovation." These claims may be investigated in the social sciences, and if they were disconfirmed, it would leave the values of technological progress without a sound basis, unless a new set of presuppositions were devised to replace the disconfirmed ones. Equally, their confirmation would strengthen commitment to the values. Here again, other things being equal, empirically confirmed claims imply making certain value judgments.

Criteria of Appraisal of Scientific Understanding Include Value Commitments

Unlike the first three, the fourth criticism—often associated with feminist philosophers of science—disputes that the criteria used in evaluating scientific knowledge are free from value commitments and that they can be clearly distinguished and separated from ethical/social values. This criticism draws, in part, on case studies that show that, as a matter of fact, there are scientific theories (not only in the social sciences) that have been accepted on the basis of criteria that depend on particular ethical/social values. Then, inevitably the theories—for example, about alleged genetically based differences in the mathematical abilities of males and females—contribute to supporting certain values and undermining others and to providing backing for projects that embody these values and, normally, not others. Criticism from such case studies cannot be conclusive, however, for it is open to the rejoinder that claims accepted on the basis of such value-laden criteria are not really instances of scientific knowledge. Some actually accepted theories may not be properly accepted, that is, on the basis of criteria that are not implicated in value commitments; actually accepted theories may not accord with neutrality, but that does not mean that scientific knowledge does not accord with it. To meet this rejoinder, the criticism may be completed by making this more far-reaching claim: It is not just that some theories are actually accepted partly on the basis of value-laden criteria but that empirical data combined with the proposed purely cognitive criteria always underdetermine what theories should be accepted, and so the role of ethical/social value-laden criteria cannot be avoided, even in principle.

General agreement has not been reached on the soundness of these and other criticisms of the neutrality of the social sciences and, thus, on the merits of the claim of the value neutrality of science.

Concluding Remarks

The idea that scientific knowledge neither supports nor undermines any value judgment has had a strong appeal throughout the modern scientific tradition, and despite the abundance of criticism, that appeal remains. Part of the appeal comes from what has been thought to be a consequence of neutrality, namely, that science may be considered to be part of the common patrimony of humanity; in principle, items of scientific knowledge may be used, more or less evenhandedly, by the adherents of any viable value outlook to inform their practical projects, even if in actual fact some items of scientific knowledge have special relevance for privileged outlooks. However, it follows from the criticisms that not all items of scientific knowledge may be used, in principle, to serve any viable value outlook. Nevertheless, the criticisms do not rule out that each viable value outlook might be strengthened by some items from the totality of scientific knowledge and that the totality (though not each individual item) of scientific knowledge may serve viable value outlooks evenhandedly—provided that science (including the social sciences) as a whole, as a worldwide practice, permits space for conducting research under a sufficient variety of theoretical frameworks, to ensure that the values "secreted" in the frameworks adopted in research are not limited to a few specially privileged ones. Then, even if the criticisms are sustained, the proposal that "scientific knowledge as a whole, the totality of scientific knowledge and properly accepted theories, serves evenhandedly all viable value outlooks," which retains much of the appeal of the idea of the neutrality of science, might still be defended as a viable ideal.

Hugh Lacey

See also Feminist Critiques of Social Science Applications; Feminist Epistemology; Ideology; Objectivity; Realism and Anti-Realism in the Social Sciences; Relativism in Scientific Theories; Science and Ideology; Technoscience and Society; Weber and Social Science: Methodological Precepts

Further Readings

Bhaskar, R. (1986). *Scientific realism and human emancipation.* London, England: Verso.

Lacey, H. (1999). *Is science value free? Values and scientific understanding.* London, England: Routledge.

Longino, H. E. (1990). *Science as social knowledge*. Princeton, NJ: Princeton University Press.
Proctor, R. N. (1991). *Value-free science?* Cambridge, MA: Harvard University Press.
Putnam, H. (2002). *The fact/value dichotomy and other essays*. Cambridge, MA: Harvard University Press.
Root, M. (1993). *Philosophy of social science*. Oxford, England: Blackwell.
Taylor, C. (1985). Neutrality in political science. In *Philosophical papers 2* (pp. 58–90). Cambridge, England: Cambridge University Press.
Weber, M. (1949). The meaning of "ethical neutrality" in sociology and economics (E. A. Shils & H. A. Finch, Eds. & Trans.). In *The methodology of the social sciences* (pp. 1–48). New York, NY: Free Press.

VERIFICATIONISM

Verificationism holds that a meaningful statement or a legitimate belief is one that is verifiable by experience. It has its roots in the empiricism of David Hume (1711–1776) and the positivism of August Comte (1798–1857), and it branches out into classical American pragmatism and logical positivism. It is often thought to be a kind of scientism that rules out, as spurious or meaningless, metaphysics, theology, ethics, and other "soft" areas of inquiry.

But while verificationism might be the enemy of some kinds of metaphysics and theology, matters are less straightforward with respect to social and moral philosophy. Verificationists have always been alert to the consequences for non–bench-based areas of inquiry. Hume had called for a science of human nature or, as the subtitle of his *A Treatise of Human Nature* has it, *An Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*. Comte's *Cours de Philosophie Positive* argued that if we are to prevent the disintegration of society, we must extend the method of science to social and moral thought. Indeed, he is the founder of the discipline of sociology.

When the early American pragmatists—Chauncey Wright, William James, and C. S. Peirce—picked up these verificationist ideas, they too took "science" to be very broad based, including not just the physical and biological sciences but also political economy, economics, and political science. The pragmatist version of verificationism argues that experience goes beyond what our five senses deliver, that we have experience in mathematical contexts, that we might experience value, and that all inquiry must and can be thought of as being a part of a seamless whole.

Verificationism these days is usually taken to refer to the core doctrine of logical positivism—the position that arose in the mid-1920s in Vienna and Berlin and that was imported to America and England at the outset of the World War II. Their aim was to unify all inquiry under the umbrella of science. The "verifiability principle" did most of the work. It was a semantic doctrine, holding that all meaningful sentences are reducible, via formal deductive logic, to statements that are empirically verifiable. Domains of inquiry can achieve clarity and progress by having their theories symbolized in the ideal and clear language of logic and cashed out in observation.

For instance, Rudolf Carnap's version of the verifiability criterion had it that deductive axiomatic theories are given empirical meaning by definitions that hook up the primitive terms in the formal language with observables in the world. He tried to show how we could give precise definitions of all scientific terms, definitions that bottomed out in a primitive language—the "thing-language" or the language of physics.

Philosophy is to get with the program, put its theories in scientific language, and render itself clear. Most of the age-old questions and their purported answers will be shown to be fruitless and meaningless, as they are not reducible to observation statements. They are not empirically verifiable, and so they are *pseudopropositions*. Statements about essences, the Absolute, the thing-in-itself, and so on are quite literally meaningless.

Ethics is in almost as precarious a state. It is either to be reimagined or imperiled. Statements about what is right or wrong are (a) statements about what people actually approve of, not what they ought to approve of—that is, ethics is an empirical science; (b) statements that express emotions or feelings; or (c) meaningless statements. As an example of the second view, we find some of the logical empiricists advocating the "Boo-Hurrah" theory of ethics, in which to say that an act is odious is to say "Boo-hiss!" to it and to say that an act is good is to say "Hurrah!"

Many of the logical positivists, however, opted for one or other version of (c): Ethics and other non–bench-based subjects are not within the bounds of science and hence are spurious. Others, Otto Neurath, for instance, argued for (a), that the social and human world is within the domain of empirical science.

The verifiability principle came under sustained pressure. It faced some formidable objections and was constantly undergoing revision and liberalization in light of them. One set of objections centers around the strength of the verifiability required. If a meaningful statement is one that can conclusively be shown to be true or false, then all kinds of discourses are in trouble. For instance, statements about the past, the future, and the mental states of others are not conclusively verifiable by observation and thus are swept away as meaningless on the strong-verifiability criterion.

Even the statement "blue, here, now," when presented with a patch of blue, is not conclusively verifiable. The perceiver might, for instance, be hallucinating or suddenly be color blind. Indeed, much of science seems to fail the test. For instance, hypotheses about unobservable entities, scientific laws, and dispositional hypotheses seem not to meet the bar. A scientific law is a universal generalization that ranges over an infinite domain, and hence, no finite number of positive instances will conclusively verify it. Statements containing dispositional terms such as *soluble* or *temperature* are analyzable only by counterfactual or subjunctive conditionals: "Were *x* to be placed in water, then it would dissolve" or "Were a thermometer to be in contact with *x,* it would register *y* degrees." Neither kind of statement is subject to conclusive verification.

Many moves were made in an attempt to overcome these challenges. The verifiability principle underwent liberalizations such as not requiring conclusive verifiability, not taking verifiability to be the entirety of meaningfulness, extending deductive logic with inductive logic, and so on. Each move took the verifiability principle farther away from the goals of straightforwardness, clarity, rigor, precision, and certainty, which drove the reductionist program of analyzing meaningful sentences via logic and observational predicates.

As a result, no one wants to be called a verificationist these days. Nonetheless, plenty of positions of the more moderate pragmatist stripe remain, and these positions may carry with them real insights for social and moral philosophy.

Cheryl Misak

See also Empiricism; Logical Positivism/Logical Empiricism; Metaphysics and Science; Positivism, History of; Pragmatism; Scientific Method

Further Readings

Ayer, A. J. (1936). *Language, truth and logic.* New York, NY: Oxford University Press.

Carnap, R. (1968). *The logical structure of the world.* London, England: Routledge.

Misak, C. (1995). *Verificationism: Its history and prospects.* London, England: Routledge.

Neurath, O. (1983). Sociology in the framework of physicalism. In R. S. Cohen & M. Neurath (Eds.), *Philosophical papers, 1913–1946* (pp. 58–90). Dordrecht, Netherlands: Reidel.

Wright, C. (1971). *Philosophical discussions* (C. E. Norton, Ed.). New York, NY: Burt Franklin. (Original work published 1877)

Vico's *Scienza Nuova*

Giambattista Vico (1668–1744) was born in Naples and lived there throughout his life. He was professor of Latin Eloquence at the University of Naples from 1699 to 1741 and was named Royal Historiographer by Charles of Bourbon in 1734. His major work, *Scienza nuova* (*New Science*) has the full title of *Principles of New Science of Giambattista Vico Concerning the Common Nature of the Nations* (1730/1744). In it, Vico puts forth a genetic account of what he terms the "great city of the human race," the basis for the generally held view that he is the founder of the philosophy of history. This genetic approach to the comprehension of human society is based on a conception of natural law that Vico opposes to 17th-century natural law theory, especially that of Hugo Grotius, Thomas Hobbes, Samuel Pufendorf, and John Selden.

These thinkers have a nondevelopmental conception of society, in which society originates or can be regarded as originating through a covenant formed among peoples such that they are transformed, as Hobbes describes it, from a state of war of all against all, in which life is "nasty, brutish, and short," into a state of governance. Vico calls the view of natural law held by such thinkers as "the natural law of the philosophers." Natural law or natural right in this sense is an ideal posited in thought to which systems of law are to respond and against which they can be judged.

Against this ideal sense of natural law, Vico opposes a conception of "universal law" based on the *ius gentium,* or "law of the peoples or nations," in Roman law. *Ius gentium* in Roman law is that

part of Roman law that is common to all other systems of law and is distinct from *ius civile,* or "positive law," law as enacted by a particular authority within a particular society and varying from society to society. *Ius gentium* is that part of law that is actually held in common by all nations and that makes law, law in each system of laws.

Vico reconceives *ius gentium* from a static principle common to all nations to a dynamic principle that he calls "ideal eternal history" and that refers to a pattern of development common to all nations. All nations develop in terms of a course of three ages: (1) an age in which society is formed in terms of gods; (2) an age of heroes, in which society is dominated by heroic figures who embody the virtues necessary for human custom; and (3) an age of humans, in which society is governed by written law and rationalistic thought. Gods and heroes are formed by what Vico calls imaginative universals, not the abstract universals of reason. With the loss of the agency of gods and heroes, the secular mentality of the third age dissolves into a rational madness that Vico calls "the barbarism of reflection." The nation returns to the necessities that govern primitive life, and a recourse of the three ages can ensue. Vico's reading of Western history regards the age of gods and heroes to have occurred before Homer and the age of humans to commence with the arrival of Greek science and philosophy, ending in the fall of Rome. The recourse begins in a return to religion, developing into the heroic society of the high Middle Ages that culminates with Dante, followed by the inception of Renaissance philosophy and culture that extends into Vico's own time and ours.

Vico claims that because humans make history, they can in turn make a science of history. This claim is based on Vico's principle that "the true is the made." We can make a science of the world of nations because the true and the made are convertible. Thus, there can be a science (*scienza*) of history, but there can be only a consciousness (*coscienza*) of nature. Since natural objects are not made by us, "natural science" can produce only a kind of exact awareness, not a complete knowledge or truth of them.

In his philosophy of history, Vico separates sacred history from the history of the gentile nations. The gentile nations arise from the offspring of the sons of Noah after the world dries out from the universal flood. Only the gentile nations are governed by the ages of the ideal eternal history. The offspring of the sons of Noah become giants, many of whom are transformed into the fathers of the first families by responding to the new experience of lightning and thunder as the great forest of the world dries out. They imitate the sound of thunder and utter the first word—*Jove*—and begin to take the auspices of Jove's action in the sky. Others who do not have this response remain feral but later seek the protection of the first families and become *famuli,* or "persons in servitude," to them. In this way, social classes are originally formed that later generate the classes of nobles and plebeians.

Each nation rises, matures, and falls within history. No nation, in Vico's view, ever masters history. Yet all in the known world of nations can be studied by beginning with its origin and narrating its development. This narrative sense of historical and social science depends on what Vico calls "a new critical art," which is accomplished by a philological-philosophical method in which an empirical knowledge of the customs, languages, laws, and deeds of a people is joined with the philosophical analysis of the universal elements of human experience.

Donald Phillip Verene

See also Enlightenment, Critique of; Historicism; Hobbes's Philosophical Method: Nature–Man–Society; Narrative in Historical Explanation; Philosophy of History

Further Readings

Bayer, T. I., & Verene, D. P. (Eds.). (2009). *Giambattista Vico: Keys to the new science.* Ithaca, NY: Cornell University Press.

Verene, D. P. (2003). *Knowledge of things human and divine: Vico's New Science and Finnegan's Wake.* New Haven, CT: Yale University Press.

Vico, G. B. (1983). *The autobiography of Giambattista Vico* (M. H. Fisch & T. G. Bergin, Trans.). Ithaca, NY: Cornell University Press.

Vico, G. B. (1984). *The new science of Giambattista Vico* (T. G. Bergin & M. H. Fisch, Trans.). Ithaca, NY: Cornell University Press.

Virtual Environments and Social Interaction

The entry discusses philosophical issues raised by the emerging advanced technologies creating virtual environments in which a kind of "social" interaction between people may be said to take place. This also raises issues for social science as well as ethics.

Virtual environments have given rise to considerable philosophical speculation from the time the relevant technology first appeared: Is it possible to distinguish between the real and the virtual? What is the relation between real persons and online characters (or avatars)? Can interactions in virtual spaces lead to new ways for people to interact, perhaps even forming whole new societies?

On closer inspection, these questions depend on how virtual environments are defined. One definition that can be said to reflect a consensus among researchers in the field revolves around the notion of *presence,* the idea that people who use virtual reality technology experience being in another place or space. If it is added that this experience must be sensory (rather than, say, the imagined experience of another place that a novel or movie might provide) and that the user must also be able to interact with or navigate within the space, then this definition delimits virtual environments in the way researchers have come to use the term and also in a way that clearly sets them apart from other technologies and experiences.

This definition is not limited to "high-tech" types of virtual environments, such as those experienced by donning a head-mounted display or standing in a Cave-type display and flying around with a 3D joystick, which are also known as *immersive environments*. Rather, it includes environments like *Second Life* or popular online games like *World of Warcraft,* where thousands of avatars can interact and play with each other. The latter are also places or spaces that people experience in terms of *presence* and that they can interact with and navigate through—though in the case of games, they are engaged in highly structured role-play (which arguably diminishes the sense of presence). In any event, it is important to distinguish between virtual environments for single users and those where avatar representations can work and play together—which have become known as shared or multiuser or collaborative virtual environments. When people experience the latter or when they engage in social interaction, the term *copresence* is used.

While the technology exists for virtual touch, smell, and taste (three of the five senses), in practice most virtual environments are limited to visual and aural environments (the two other senses), though these environments also include a kind of pseudo-touch whereby objects can be manipulated (a house can be built or a ball thrown—though they are weightless!). Thus, the interaction with the environment is mainly via seeing it, hearing it, and navigating through it and manipulating objects. This also applies in multiuser environments to interacting with others' avatars. Once this is recognized, again, many of the philosophical conundrums (can people be hurt in virtual environments?) quickly disappear (in a purely visual and aural environment, not physically!).

Similarly, can avatars be used to explore multiple personalities? Yes, although this often takes the form of an exploration of visual representations of the self and how people represent themselves to other avatars. Of course, people can come to identify with one or more of these visual representations, but as in real-world encounters, it is hard to "stage" oneself in a role that does not truly present who one is. In this respect, much also depends on the type of technology used. Consider, for example, the difference in possibilities for self-representation in text-based virtual environments (many environments like Second Life have been and still are mainly text based), as against environments where voice communication is used (Second Life has also enabled communication via voice, though this is still not very common): The possibilities for misrepresenting oneself are far more restricted in an environment with voice.

Virtual environments—and especially multiuser ones—have become popular means of socializing and pursuing leisure and various other activities (e.g., education) in online spaces. They have also become useful laboratories for doing social science research. This is because they allow all interactions to be captured digitally and analyzed and provide an environment whose parameters can be altered at will and that does not have the same risks as physical environments. Examples here include environments for simulating fire escapes or for shy people to introduce themselves to others, or environments in which people can adopt different kinds of appearance to see how this affects their encounters with others. Many interesting and useful insights about social interaction have been produced using virtual environments in this way.

Yet these research possibilities also present some new conundrums in terms of the *ethics* of research: How far can researchers go in exposing research subjects to harm, even if the harm experienced is mainly psychological? Environments that are patently not real are experienced as such, which makes some experiments in these environments problematic. Another issue is raised by the fact that some environments are rich with data—but almost like a "big brother" society under surveillance. Here, we can think of Second Life or online games, which, although they are public, allow researchers

to watch and analyze behaviors in detail and in a perfectly panoptic way. Arguably, while this may be legal, it may also be intrusive—or worse—from the viewpoint of research ethics.

This entry has confined itself to a tightly defined idea of virtual environments and discussed only a few related issues and prospects for the social sciences. If a broader view is taken whereby *virtual* is taken to mean any online phenomena, then the net could be cast much wider too in terms of the social implications. What about the virtual relationships people have via social networking sites such as Facebook? Or the addiction to gambling online? Or the followings that celebrities have on Twitter? In this wider picture, much of the knowledge and information we access, the relationships we maintain, and the way we entertain are moving into the digital realm. This is a much broader sense of *virtual,* and yet the narrower notion of virtual environments discussed here can shed much light on these more general social issues—since interaction with immersive virtual spaces can be seen as an extreme version of less immersive and more quotidian experiences via the Internet and the Web.

Ralph Schroeder

See also Experiments in the Social Sciences: Ethical Issues; Human–Machine Interaction; Information Society; Technoscience and Society

Further Readings

Blascovich, J., & Bailenson, J. (2011). *Infinite reality: Avatars, eternal life, new worlds and the dawn of the virtual revolution.* New York, NY: HarperCollins.

Schroeder, R. (2010). *Being there together: Social interaction in virtual environments.* New York, NY: Oxford University Press.

Schroeder, R., & Bailenson, J. (2008). *Research uses of multi-user virtual environments.* In R. Lee, N. Fielding, & G. Blank (Eds), *The SAGE handbook of online research* (pp. 327–342). London, England: Sage.

Virtue Epistemology

Virtue epistemology is a distinctive approach to understanding the evaluative and metaphysical dimensions of cognition. Although the field is marked by considerable diversity and disagreement, there is broad agreement on at least two fundamental principles. On the one hand, virtue epistemologists agree that cognition is *normative*. Cognitive science has much to teach us about how we perceive, remember, reason, inquire, and so on, but unfortunately, there is no easy path from these extremely valuable empirical insights to conclusions about how we *ought* to cognize or what counts as *good* cognition. This is not to say that empirical facts about cognition are irrelevant to the normative questions but only that important questions remain once all the science is in. On the other hand, virtue epistemologists agree that the ultimate source of epistemic normativity, and hence the central focus of epistemological inquiry, are cognitive agents and communities, along with the fundamental powers, traits, and habits that constitute their intellect. This contrasts with the mainstream approach in the analytic philosophy of the later 20th century, which focuses on individual beliefs and inferences instead of individuals and their cognitive character. Traits that promote good cognition or intellectual flourishing are called, following a tradition extending back to Aristotle, *intellectual* or *cognitive* or *epistemic virtues,* hence the name "virtue epistemology."

Virtue epistemologists try to answer long-standing philosophical questions about cognition by focusing on how an agent's intellectual powers, habits, and abilities ("dispositions" for short) enter into the conduct of inquiry and formation of belief. A crucial resource here is the notion of an outcome *manifesting* a disposition, which is an especially intimate relationship between the outcome and the disposition. For example, one important philosophical question about cognition is "When is a belief based on perceptual experience?" A virtue epistemologist might answer, "When the perceptual experience causes the subject to form the belief, and the fact that it does so manifests the subject's disposition to trust his senses." Another question—perhaps the most important question in this area—is "What is knowledge?" A virtue epistemologist would answer, "Knowledge is true belief manifesting epistemic virtue." Another important question is "Why is knowledge more valuable than true belief?" A popular answer among virtue epistemologists is "Because you do not necessarily deserve credit for believing the truth, as might happen if you luckily guessed the correct answer to a question; by contrast, you know something only if you deserve credit for arriving at the truth, through

the exercise of your epistemic virtues, which makes knowledge better than mere true belief."

What counts as an epistemic virtue? A standard answer is that one central and important class of virtues includes the subject's truth-conducively reliable doxastic dispositions. That is, they are dispositions that make the subject good at detecting and endorsing the truth, so that he or she usually gets it right when he or she exercises those dispositions. A further set of important and, from a philosophical perspective, poorly understood dispositions concerns the metacognitive task of suspending judgment on a question. It is not obvious that the quality of these dispositions can be measured simply in terms of how reliably they produce true beliefs, since suspending judgment occurs only if one *refrains* from forming a belief. So when ought one to suspend judgment? Most, if not all, of our cognitive dispositions have innate biological and social bases, so the abstract account of epistemic virtue put forward by the virtue epistemologist must be properly supplemented through insights from the biological, cognitive, and social sciences.

The nature and scope of epistemic virtues is an area of potentially fertile interdisciplinary work among philosophers and scientists. For example, some virtue epistemologists, often called "virtue responsibilists," accept a more demanding conception of epistemic virtue than the minimal reliabilist conception of virtue mentioned above. Responsibilists define epistemic virtues as praiseworthy and refined character traits with a distinctive motivational profile, such as conscientiousness and open-mindedness, which underwrite robust, broad-based dispositions to inquire well across a wide range of circumstances. These epistemic traits share the same profile as the ethical traits featured in the Aristotelian tradition of virtue ethics, such as generosity, justice, and compassion. But a rich body of work in social psychology has led many psychologists and philosophers alike to question the existence of such traits. For instance, in one set of experiments involving seminarians at Princeton Seminary, the strongest predictor of whether a seminarian would stop to help a stricken person was how much time the seminarian thought he had to arrive at a lecture on time, not on how compassionate he was. In another set of experiments, whether the subject offered help to a distressed passerby was strongly influenced by whether the subject had recently found a dime in a payphone, not by how compassionate she was. These and other equally surprising results are often taken to suggest that we dramatically overemphasize the prevalence of individual traits in determining behavior and correspondingly underemphasize the impact of situational factors. If this is correct, then it does not bode well for the responsibilist conception of epistemic virtue. Does the hypothesis that we have refined epistemic virtues fare any better than the hypothesis that we have refined ethical virtues? This is an issue ripe for further experimental work, which would have important philosophical consequences.

The cultivation of epistemic virtues is another area of fruitful overlap between virtue epistemology and the sciences. Just as there is no doubt that a human's biological endowment heavily influences his or her cognitive character, there is equally little doubt that his or her socialization has a similar effect. For example, consider how much humans rely on testimony. Competently consuming testimony involves a battery of skills and presuppositions. There is strong evidence that a speaker's social status, gender, and ethnicity affect how his or her testimony is regarded. Unless we are disposed to be appropriately sensitive to features relevant to the quality of testimony, and appropriately insensitive to irrelevant features, we run the risk of incompetently or unfairly consuming testimony. Scientific research is essential to help us understand our habits and predispositions, both vicious and virtuous, in this regard.

John Turri and Ernest Sosa

See also Epistemology; Experiments in Social Science; Feminist Epistemology; Normativity; Social Epistemology

Further Readings

Baehr, J. (2011). *The inquiring mind.* Oxford, England: Oxford University Press.

Doris J. (2002). *Lack of character.* Cambridge, England: Cambridge University Press.

Fricker, M. (2007). *Epistemic injustice.* Oxford, England: Oxford University Press.

Greco, J. (2010). *Achieving knowledge.* Cambridge, England: Cambridge University Press.

Greco, J., & Turri, J. (Eds.). (2012). *Virtue epistemology: Contemporary readings.* Cambridge: MIT Press.

Pritchard, D. (in press). Anti-luck virtue epistemology. *Journal of Philosophy.*

Sosa, E. (2007–2009). *Apt belief and reflective knowledge* (Vols. 1 and 2). Oxford, England: Oxford University Press.

Sosa, E. (2011). *Knowing full well.* Princeton, NJ: Princeton University Press.

Turri, J. (2011). Manifest failure: The Gettier problem solved. *Philosophers' Imprint, 11,* 1–11.

Zagzebski, L. (1996). *Virtues of the mind.* Cambridge, England: Cambridge University Press

Zagzebski, L. (2009). *On epistemology.* Belmont, CA: Wadsworth.

Weber and Social Science: Methodological Precepts

Max Weber's views on social-scientific methodology are rather multifaceted and complex. Efforts by scholars to present them entail the risk of oversimplification or of emphasizing certain aspects of his thought while neglecting others. This entry tries to avoid such distortions by discussing Weber's methodological precepts with attention both to the influences on his work and to the influences his work has exerted on others.

Introduction

Weber's views on social-scientific method elude easy characterization. He has been construed as an antipositivist by scholars pointing to his famous notion of *Verstehen,* while, at the same time, he has also been taken to be a kind of positivist mainly because of his adherence to a strict distinction between "is" and "ought"—that is, facts versus values in social-scientific reasoning. Both these characterizations are exaggerated, since Weber's *Verstehen* method for understanding social action must be seen as what it really is, namely, a rational process—not one based on empathy—while his view on the fact/value distinction is less blunt than it appears in his popular "twin lectures" of 1917 and 1919. Moreover, in his methodological works, value freedom as an ideal for scientific research is qualified as value relation, thus bridging the gulf between "is" and "ought."

According to Paul Honigsheim, Weber's biographer, Weber hated the sharpening of methodological knives if there was nothing on the table to be carved. Nevertheless, his methodology is more often extracted from his methodological essays rather than from its applications. Weber was not a "card-carrying" philosopher, and he himself admitted that he was not philosophically minded. Yet he was never a sociologist by profession either, except for a part-time assignment during the last year of his life as professor of political economy in Munich. His first teaching position was in commercial law. Insofar as methodological issues were concerned, he relied on others—for example, his Heidelberg colleague, the Neo-Kantian philosopher Heinrich Rickert. Weber's references to Rickert are quite explicit. The relationship between the two thinkers, however, is problematic. The use of Rickert might have been a matter of convenience and does not give us the full picture of the influences on Weber's thought. Systematizing Weber along Rickertian lines would neglect the explicit reservations in Weber's views on Rickert's terminology, such as in the so-called Nervi fragment, a small note from Max to Marianne Weber from 1903 (extensively discussed by Bruun).

Weber is best seen as a synthesizer who borrows from others: For example, the concept of ideal type is taken over from Georg Jellinek, though it gets transformed. To Weber, concepts are human constructs—in contrast to the Hegelian tradition. Weber's Neo-Kantian nominalist concept formation procedure allows for an intersubjective selection from a vast and chaotic reality, based on postulated

points of departure. Intersubjectivity is a precondition for cumulativity in the process of knowledge production. For the historical school, by contrast, abstraction from reality presented a major problem they could not account for in their methodological controversy with Austrian marginalist economics. However, Weber does not exhaust the topic of how concepts are actually formed.

Perspectivism is sometimes a term used to catch the essence of Weber's methodology, in which value hierarchies are rationalized from a postulated point of departure ("interest"), so the resulting policy recommendations can be scrutinized regarding their logical consistency.

In *Gesammelte Aufsätze zur Religionssoziologie* (*GARS;* see Further Readings), at least in the section Intermediate Reflections, Weber expresses explicitly a methodological creed coherent with several other texts and, in fact, deals with the notion of ideal type as well as with what can be seen as the unique Western scientific creed. However, Weber's methodology is not only to be found in his explicit methodological texts. It can also be found in its applications. Already in the 1890s, Weber applies the methodology he later develops. *Freiburger Antrittsrede* (The National State and Economic Policy, 1895/1994) is a good example when explaining the logic of the agrarian policy east of the Elbe and why German settlers are replaced by Polish migrant workers. It is pioneering policy science—rationalizing value hierarchies for the purposes of an instrumental decision making.

There is no single Weberian school of devoted followers; instead, we have competing interpretations. Weber's most famous methodological tool, the so-called ideal type, has been—in very different ways—a starting point for the three main methodological paradigms in social science, namely, Parsonsian macrosociological functionalism, Schützean hermeneutics, and Lazarsfeld's empirical survey tradition (we shall return to these later). The ideal type is a heuristic tool separating hypothesis from theory. There are several ideal types of social entities or social phenomena. The one regarding bureaucracy is perhaps the most famous, closely followed in importance by the historical types of Protestant ethic and the spirit of capitalism. The bureaucracy ideal type, however, is not an *ideal* ideal type, since it is a general type rather than a historical individual, such as the ones we find in *GARS*. One encounters bureaucracy in many historical locations, from ancient Egypt to the Soviet Union. However, the growth of modern bureaucracy is part of an irreversible rationalization process, which is at the core of Weber's philosophy of history.

Weber's Methodology

Weber's methodology could be encapsulated in the formula "three controversies, two traditions, and one predicament."

a. The *predicament* refers to the post-Enlightenment dilemma of an anxiety of choice—in other words what Weber's neighbor Ernst Troeltsch labeled the "polytheism of values." Weber's methodology in social science corresponds to a Kierkegaardian-Nietzschean predicament in philosophy. The criticism voiced would go like this: By postulating a top value, we can rationalize value hierarchies in the service of instrumental social policy. There is no scientific way to judge between competing ultimate values, such as liberty or equality.
b. The historical school, dominant at the time, was unable to account for the principles of selection behind their identification of central forces in history. Historicism was ever present in Germany at that time, and it was idiographic rather than nomothetic.

The *two traditions*—in addition to the prevailing historicism just mentioned—were Neo-Kantianism (in Weber's case, the Neo-Kantianism of Rickert's southwest or "Baden" school) and Austrian political economy (i.e., Carl Menger's marginalism). Both of them left their imprint on Weber's "value–aspect–choice methodology," in which interests or perspectives played the central role in explaining how social-scientific research and the choice of themes could be governed by the interests or perspectives of the researchers involved.

The *three controversies* were as follows. Weber's ideal type, is an attempt to help the historicists in the Schmoller camp (the younger historical school) out of their cornered position in the famous *Methodenstreit,* the central controversy forming the background to Weber's methodology. The other two controversies were the *demarcation or specificity debate* and the *objectivity debate*. Let us discuss these in turn.

1. The *Methodenstreit* (controversy over method) between Gustav Schmoller and Carl Menger peaked in the 1880s. The central issue was how

to establish laws. This placed the historicists in a vulnerable position, since their natural inclination to the contrary was to be ever more idiographic, in the spirit of Leopold von Ranke. So by wishing to be idiographic, they were handicapped in terms of laying down the laws of history. Throughout the 19th century, history (seen as idiographic) versus theory (seen as nomothetic) was the main divide, and it still is today. Weber launched his ideal-type method in response to the *Methodenstreit*.

2. The *demarcation debate* or *specificity controversy* had to do with a discussion carried out within the antipositivist–historicist camp regarding the grounds on which the specificity of cultural or historical "science" (i.e., the social sciences) could be argued. The aim was to carve out a safe realm for the human or social sciences that could not be easily invaded by the ideal of unified science. In this way, the social sciences were to be demarcated from the natural sciences by the use of a certain criterion: Were the two kinds of science different in terms of content or the methodology employed? In the specificity debate, Weber actually sides with Wilhelm Dilthey, the father of hermeneutic understanding, in that the *object* (and not the method) is the main difference between positivist natural-scientific explanation and the "moral" sciences. But Weber chose to do so by applying and developing Rickert's method, with its lucid concept formation, by explicit value premises, as Gunnar Myrdal later would call them. Yet it is rather a myth that Weber is a close follower of Dilthey. Weber refers to Dilthey 14 times; however, almost all of them are in the first Roscher and Knies essay (1903), the first in a series of methodological essays from 1903 to 1907. Weber never uses the term *hermeneutics*. His ideal types of rational action are formulated so that "even a Chinese would understand."
3. The *controversy over objectivity:* Around the turn of the 20th century, the problem of intersubjective confirmation and objectivity was addressed full-front in social science, although philosophers such as David Hume and Jeremy Bentham had paved the way a long time before.

The objectivity debate, regarding testable and intersubjective truth, still goes on, although it peaked within the German Social Policy Association in 1913. The interpretations of Weber's position on offer are debatable, but very few would deny that he is an anti–natural law advocate, something made explicit in the early paragraphs of *Economy and Society.* In trying to understand his position on objectivity, we must keep in mind that in Weberian anti-metaphysics there are no firm starting points and no transcendental truths. Instead, a point of view has to be postulated as a starting point or goal value posited for policy recommendations. Going back to what was mentioned earlier, this ultimate value cannot be proved by means of science. Whether such starting points might be found by other means is a somewhat open question in Weber's case. There is, moreover, a long debate about Weber's own ultimate value, whether it was the contemporary cultural values of liberalism or nationalism. However, in "Suffrage and Democracy" one gets the sense that culture itself, more precisely German culture, is the ultimate value.

In the famous "Objectivity" essay of 1904, Weber's notion of ideal type is formulated in a negative way, as a list of what it is *not*. He uses the term *einseitige Steigerung* (one-sided accentuation), indicating that concept and reality are not isomorphic. *Steigerung* has often been supposed to be central for the understanding of the ideal type as heuristics. It is, however, only used once in Weber's collected methodological writings. It is further often assumed that Weber's ideal type is a method of empathy; this is an exaggeration. Weber's interpretive *Verstehen* is of a rational kind, and Carl Menger's rational actor model served as a prototype or model for him. Weber's *Verstehen* has nothing to do with psychology, neither in the sense of a "basic law of psychophysics" nor in a hermeneutic empathetic sense. This can be seen in the use of the card game Skat as an example in his methodological essays, in which understanding is an act of rational interpretation followed by ad hoc explanations of anomalies.

Additional Influences

The account offered so far does not exhaust all the influences on Weber's methodology.

a. John Stuart Mill's canons of comparative research were very likely an additional tacit influence on Weber's comparative sociology of religion. Weber, in a letter to the medieval historian Georg von Below, explains why he needs to underpin his Calvinist thesis laid out in *The Protestant Ethic and the Spirit of Capitalism*

by a study of all other rational religions. For a German scholar, British influences could be a rather sensitive issue at the time, despite *Geisteswissenschaften* being a translation of Mill's "moral science."

b. Weber also refers to Marx's analysis of the logic of the capitalist mode of production as an ideal type. Weber's historical materialist influences are present in *The Social Causes of the Decline of Ancient Civilization* (1896), in contrast to his Calvinist thesis.

c. Yet another less visible influence is Georg Simmel. They knew each other well and took each other for granted, without much need for explicit references. Weber's copy of Simmel's book—kept in the Weber archives in Munich—on Nietzsche and Schopenhauer is full of Max Weber's annotations in the margins.

Weber's Influence on Contemporary Social Thought

Weber's methodology has been extremely influential, serving as a starting point for several extensions. Below, we shall present the three prominent cases we introduced earlier: (1) Talcott Parsons's macrosociological functionalist system, (2) Paul Lazarsfeld's survey tradition, and (3) Alfred Schütz's phenomenology.

But first some different examples. (a) A noteworthy exception where no influence can be seen is that of the Chicago school, despite both its historical and its thematic affinities with Weber's thought. (b) Another case is the Frankfurt school, which appears to be a synthesis of Weber and Marx. Jürgen Habermas wrote a whole book on Weber, which, however, has played a rather minor role in modern Weberology. (c) On a different front, as far as rational choice theory goes, there is again not much of an influence. Employing Weber as a rational choice pioneer has its limits, since his research deals mainly with unintended consequences in need of idiographic clarification. (d) On the other hand, Weber's so-called scientific value relativism has been particularly influential in Scandinavia, with its strong anti-metaphysics inclination. It was initiated by Axel Hägerström's "value nihilism" and then carried on within Scandinavian legal realism by scholars such as Gunnar Myrdal and Alf Ross.

Parsons's misrepresentation of Weber is rather obvious; already in the headline of his translation of *Wirtschaft und Gesellschaft* (*WuG; Economy and Society,* Part 1), Parsons used the label *theory,* although Weber explicitly states that it is *not* a theory. However, Parsons might well have saved Weber's sociological legacy for posterity by producing a macrosociology out of his taxonomy in *Economy and Society*—a successful wishful extension that became a cornerstone of sociology.

Lazarsfeld also displays a paradigmatically biased reaction to the ideal type, denouncing it rather as an inhibiting mishap and negative test case suffused by the German "*Verstehen* language." Lazarsfeld instead lists a number of surveys carried out by Weber as pioneering, such as those in his investigations in industrial sociology from 1908, his survey of the design of opinion polls, his survey of the press, and so on.

Schütz does not neglect the problem of intersubjectivity, one of Weber's most crucial concerns. Schutz wrestles with transcultural empathy and is stuck with a lack of rational evidence, something that Weber provides with the rational actor model.

That three dominant and competing paradigms are built from the same cornerstone is a problem insofar as they all claim to be interpretations of Weber. It illustrates an inherent problem with approaches to classic texts that read them as if they were valid in present times. Classic texts should be left in their contexts and are better understood there.

Concluding Remarks

Max Weber can be seen as an antisociologist within sociology. He is part of a longer tradition of secularization and calculability going back to as far as Niccoló Machiavelli, with Thomas Hobbes and Jeremy Bentham as midway stations. Despite his enormous influence, he is not a genuine innovative classic introducing a wholly new paradigm or a "rupture" in the Kuhnian sense. He is rather a mediator.

It is also fair to say that Weber's writing a century ago is no up-to-date guide in methodology. He is nevertheless still with us, since the main methodological and philosophical divides surveyed above have remained the same for the past two centuries. He thus occupies a prominent strategic place in intellectual history. His legacy is rich yet split into many currents.

If you build at a crossroad, you get many followers.

Sven Eliaeson

See also Explanation Versus Understanding; *Methodenstreit; Naturwissenschaften* Versus *Geisteswissenschaften;* Philosophy of Sociology, History of; Value Neutrality in Science; Weber's *Verstehende* Approach

Further Readings

Brecht, A. (1959). *Political theory: The foundations of twentieth-century political thought* (see especially chap 3 in Pt. 1). Princeton, NJ: Princeton University Press.

Bruun, H. H. (2007). *Science, values and politics in Max Weber's methodology.* London, England: Ashgate. (Original work published 1972)

Dahrendorf, R. (1987). Max Weber and modern social science. In W. J. Mommsen & J. Osterhammel (Eds.), *Max Weber and his contemporaries* (pp. 574–580). London, England: Unwin Hyman.

Eliaeson, S. (2002). *Max Weber's methodologies: Interpretation and critique.* Cambridge, England: Polity Press.

Lazarsfeld, P., & Oberschall, A. (1965). Max Weber and empirical social research. *American Sociological Review, 30,* 185–199.

Myrdal, G. (1990). *The political element in the development of economic theory.* New Brunswick, NJ: Transaction. (Original work published 1929)

Schütz, A. (1973). *The structure of the life world.* Evanston, IL: Northwestern University Press. (Original work published 1932)

Tenbruck, F. H. (1987). Max Weber and Eduard Meyer. In W. J. Mommsen & J. Osterhammel (Eds.), *Max Weber and his contemporaries* (pp. 234–267). London, England: Unwin Hyman.

Turner, S. P., & Factor, R. (1994). *Max Weber: The lawyer as social thinker.* London, England: Routledge.

Weber, M. (1971). Research strategy for the study of occupational careers and mobility patterns. In J. E. T. Eldridge (Ed.), *Max Weber: The interpretation of social reality.* London, England: Michael Joseph. (Original work published 1908)

Weber, M. (1976). The social causes of the decline of ancient civilization. In *Max Weber: The agrarian sociology of ancient civilizations* (R. I. Frank, Trans.). London, England: New Left Books. (Original work published 1896)

Weber, M. (1978). *Economy and society.* London, England: University of California Press. (Original work published 1922)

Weber, M. (1984–). *Max Weber-Gesamtausgabe* [Collected works of Max Weber]. Tübingen, Germany: J. C. B. Mohr. (Includes the following titles: *Wirtschaft und Gesellschaft* [Economy and society], original work published 1922; *Gesammelte Aufsätze zur Wissenschaftslehre* [Collected essays on epistemology], original work published 1922; *Gesammelte Aufsätze zur Religionssoziologie* [Collected essays on the sociology of religion], original work published 1920–1921; and *Gesammelte Politische Schriften* [Collected political miscellanies], original work published 1921; see http://www.maxweberstudies.org for specifics)

Weber, M. (1994). *Political writings* (P. Lassman & R. Speirs, Eds.). Cambridge, England: Cambridge University Press.

Weber, M. (1999). *Max Weber: Essays in economic sociology* (R. Swedberg, Ed.). Princeton, NJ: Princeton University Press.

Weber, M. (2012). *Max Weber: Collected methodological writings* (H. H. Bruun & S. Whimster, Eds.; H. H. Bruun, Trans.). London, England: Routledge.

Weber's *Verstehende* Approach

A *Verstehende* ("understanding," from the German *verstehen,* "to understand") approach emphasizes that any systematic empirical study of culture and society must begin and end with the examination of their meaning as experienced in the lives of human beings. Similarly, human behavior must be understood as *meaningful action*. In Max Weber's (1864–1920) case, the approach found expression in three main areas: (1) in accounts of motivation, most famously in the Protestant work ethic; (2) in reflections on the inner logic of life spheres, such as religion or the economy; and (3) in Weber's ideal types of bureaucracy or individual rational action. Underlying all the examples of his approach was the persistent demand that the social scientist should account for the course of human affairs by relating it to the intentions, beliefs, and actions of individuals.

The Person and His Time

The German intellectual context of Weber's work involved intense disputes around the application of natural-scientific methodology in fields as diverse as experimental psychology, economics, and sociology. Darwinian science and an aggressive Nietszchean secularism challenged Christian beliefs and the Kantian philosophical grounding of ethics, while Marxism denied the separation of theory and practice in its confrontation with existing political structures.

Weber came from a family connected with the textile industry and national politics. He became a lawyer and social researcher before becoming a professor of economics and finance in 1896. His fierce polemics gained him national recognition in both academic and public policy debates, asserting both the intellectual independence of academic social science and the relevance of its methods and findings to political life.

In a context of rising nationalism, Weber combined liberal individualism with patriotism and a commitment to scientific objectivity. He recognized fully the contradictions that arose between them. Even while asserting that the individual was the basic unit, the atom of social science research, he acknowledged the deep internal conflicts experienced by the modern individual and in his own life. Indeed, for him, conflict was a primary aspect of modernity, unresolvable except by retreat into old faiths that he could no longer share (his mother had been a devout Protestant Christian). His vast legacy of correspondence with friends and colleagues, as well as his published work, are the most compelling evidence, unrivalled in its extent, of the intimate connections between personal agonies, political outlook, and intellectual contributions to knowledge.

Themes

Weber left no systematic account of his *Verstehende* approach. Considering the many and varied ways it was exemplified in his work, we can group his concern for *Verstehen* under three main headings: (1) individual motivation, (2) life spheres, and (3) ideal types.

Individual Motivation

In the first place, individuals are bundles of needs, drives, passions, and beliefs. Weber did not consider himself a psychologist, although at one point he conducted a survey of psychological factors in factory workers' output. But it was always the framing of these in social relations and the way they were translated into meaningful patterns of action that was his primary focus. The most celebrated example of this theme occurred in his 1904–1905 study of the motivations behind early capitalism, *The Protestant Ethic and the Spirit of Capitalism*. We arrive at an understanding of the individual's response to the idea of predestination not from its logic but by appreciating the anxiety about salvation it promoted and the way this was allayed by hard work. In what Weber called "the paradox of consequences," a new work ethic gave an extra impetus to capitalistic behavior in this world, even though the hopes of the believer were centered on the next.

This nexus of beliefs, behavior, and economic conditions was a theme he repeated in extensive comparative studies of religion. In India and China, he found religious and ethical doctrines reinforcing enduring systems of social relations, whereas in ancient Israel, he found a persistent thread of religious rationalism demanding that everyday life should be subject to divine law and encouraging involvement in a project to save the whole people.

Very often Weber's accounts of religion have been interpreted as a direct refutation of Marxist materialism, though he rejected replacing a one-sided determinism with another that was equally one-sided. His *Protestant Ethic* essay was followed by one showing how in the United States people might join a church to gain business opportunities. He was an advocate of a multicausal interpretation of social life where ideas worked more often in indirect ways, though, drawing on the image of switching points on a railway track, he also argued they could move the underlying course of history in one direction rather than another.

Life Spheres: Coherence and Conflict

In his reflections on the underlying course of history, we can find a second theme in his *Verstehende* approach. As a German lawyer, he was familiar with the state administration and bureaucracy as a rational system of rules depending on the objectivity and sense of duty of trained officials. Later, he saw the rationality of bureaucracy as an aspect of a comprehensive process of rationalization in every sphere of life: in capitalist finance, in forms of production, and in science above all, but also in sexual behavior, where he was fully aware of trends in psychoanalysis, and in religion, where from time immemorial theologians had sought to develop ritual and dogma to justify the ways of God to human beings. However, the quest for meaning in life, which he tended to see as a universal desire under conditions of the rationalization of every life sphere, meant a growing cleavage between the search for personal fulfillment and the demands of an increasingly machine-like world, an "iron cage."

For Weber, a *life sphere* was a complex of meanings—a set of activities, ideas, and beliefs connected with each other through logic, mutual implication, and necessary association. Examples of life spheres were religion, economics, sexuality, and the media. He never attempted a systematic account of life spheres, but in their inner coherence, they had power over people and, driven to ultimate conclusions, they came into conflict with each other. The classic case was conflict between the demands for a strong state and a free market, the theme of Weber's controversial inaugural professorial lecture.

Advancing rationalization meant increasing conflict between different life spheres. Weber was the exact antithesis of a celebrator of progress and the advance of reason. Human beings pursued ideas to their logical conclusions and suffered the consequences. If they believed in complete sexual freedom, they would suffer the consequences in their personal relationships. His *Verstehende* approach conveyed a tragic view of human existence, caught between the remorseless advance of rational intelligibility in institutional life and a deepening sense of personal conflict.

Rationality and Ideal Types

Weber never addressed the nature of rationality directly, even though his work is pervaded by the terminology of rationality and he wrote at length on its influence in human affairs. He did, however, argue that in the shape of the rational ideal type, it was an indispensable instrument for social research and historical explanation. This third aspect of his *Verstehende* approach involves his most explicit concern for scientific methodology and bears his signature more than any other of his contributions to knowledge.

An ideal type is a complex of meanings that the historian or social and cultural researcher summarizes, identifying its most salient features to distinguish it from other such complexes and present it in a more clear-cut and coherent way than might ever be found in reality. It is a rational construction by the analyst rather than an existing entity. Reality and the ideal type must not be confused—not that Weber appears to have had any doubts about reality. For him, reality was what the social and cultural sciences dealt with, but the problem was that there was no way to represent reality in its fullness and it was necessary to have clear concepts to illuminate areas of investigation. Weber considered the idea of bureaucracy as an ideal type. The image of a perfectly functioning hierarchy of offices, dutiful officials, and complete records existed nowhere in reality, not even in his beloved Germany, but it served as a template against which existing state administrations could be compared.

Most famous of all his ideal types was homo economicus—economic man—the premise of classical economics, which imagined a rational actor finding the means to pursue objectives with the least possible cost, and therefore maximizing personal satisfaction, and then asking what the aggregate outcome would be in a market situation where everyone behaved in that way. Weber, as an economist who was also a historian, deemed that this kind of conduct was never fully realized in actual life but real situations approximated to it sufficiently because people strove to achieve this degree of rationality, and the resulting economic analysis was therefore a very powerful tool for understanding how real markets worked. This kind of economic theory clinched for him the worth of ideal-type analysis.

Rational action of an individual actor supplied the social scientist with the most understandable data on which to base explanations of the course of social events. When Weber sought to catalog the most elementary concepts necessary for a rigorous discipline of sociology, he identified two kinds of rationality within a fourfold classification of action. The first was a rationality of ends and means and the second, a rationality of conduct that realized values. Additionally, he recognized two types of nonrational action—the expression of emotion and the observation of tradition. These four basic types of action were the ultimate constituent elements in all his subsequent definitions of complex types of social relations, including authority, representation, organization, and the state.

Individualism in Method and in Life

Weber's typology of action exemplified a ruling dictum in his *Verstehende* approach to history and the scientific study of human affairs. All explanation had in principle to be brought back to the actions of individuals. This applied to even his most famous concept, *charismatic authority,* possessed by figures as different as, for example, the Dalai Lama and Napoleon.

It depended ultimately on individual followers acting on a belief in their extraordinary powers. Weber was correspondingly dismissive of sociologists or political scientists who sought to ground their academic discipline on collective concepts like society or the nation and denounced those who used the academic lectern to preach patriotism in their name.

Weber was both a fervent adherent to the German national cause and a dedicated believer in the rigors and discipline of scientific research. Intellectually, his position was later called *methodological individualism*. In his personal life, his individualism took the form of heroic struggle with inner demons rather than the free expression of individuality, despite the widespread popular cult of personality. The asceticism of his religious upbringing stayed with him long after he had left its beliefs behind. Weber exemplified in his life the modernity that was the focus of his work. The demand to find meaning through rationality, the quest for understanding human life on this earth when traditional answers had been rejected, was for him the fate of the modern person. The scientist took that quest to new levels and incurred the additional grief of knowing that it could never reach its goal.

Martin Albrow

See also Causes Versus Reasons in Action Explanation; Explanation Versus Understanding; Individualism, Methodological; Rationality and Social Explanation; Value Neutrality in Science; Weber and Social Science: Methodological Precepts

Further Readings

Albrow, M. (1990). *Max Weber's construction of social theory.* London, England: Macmillan.

Kalberg, S. (2012). *Max Weber's comparative-historical sociology today: Major themes, mode of causal analysis, and applications.* Farnham, England: Ashgate.

Outhwaite, W. (1975). *Understanding social life: The method called verstehen.* London, England: George Allen & Unwin.

Radkau, J. (2009). *Max Weber: A biography.* Cambridge, England: Polity Press.

Weber, M. (1978). *Economy and society: An outline of interpretative sociology* (2 vols.; G. Roth & C. Wittich, Eds.). Berkeley: University of California Press.

Weber, M. (2009). *The Protestant ethic and the spirit of capitalism with other writings on the rise of the West* (S. Kalberg, Ed.). Oxford, England: Oxford University Press.

Whimster, S. (Ed.). (2004). *The essential Weber: A reader.* London, England: Routledge.

Welfare Economics

Welfare economics is a subfield of economics; it shares many concepts with microeconomics, but it is more concerned with *improving* the welfare of economic agents than with explaining how markets operate. Welfare economics has been somewhat marginalized within the context of contemporary economics. Some authors, notably Anthony B. Atkinson, have argued that welfare economics has ceased to exist for all practical purposes. But its concepts and methods, sometimes under different guises and in revised forms, continue to be used in many settings that fall under the rubric of "normative economics." The application of economic reasoning to the problem of identifying and rectifying unfair or otherwise undesirable economic and social situations is a research project that is very much alive. It encompasses approaches that have evolved to improve on some of the dilemmas that welfare economics has stumbled on. The original concern of welfare economics with improving the welfare of economic agents (typically measured in terms of preferences over utilities, e.g., utilities derived from different levels of income, consumption, etc.) has to some extent been replaced by a concern for ways of improving their well-being defined in broader terms. This entry looks at the origins and evolution of welfare economics before concluding with a reflection on the challenges that lie ahead.

From the "Old" to the "New" Welfare Economics

Welfare economics asks, "How can societal welfare be improved?" This is not a purely hypothetical question. There are many circumstances where markets seem to fail to allocate resources adequately, such as when "negative externalities" cause some people to unfairly carry part of the costs of activities profitably undertaken by others (e.g., industrial pollution). In the mathematical language of economists, it can be expressed as follows: Maximize the function

$$W = W(w_1, w_2, \ldots, w_n),$$

subject to some feasibility constraints, in a society composed of n individuals (the welfare status of a representative individual i being represented by w_i). As long as the function remains purely generic, as stated here, the problem remains merely hypothetical. However, the pioneers of the

discipline, notably Alfred Marshall (1842–1924), F. Y. Edgeworth (1845–1926), and Arthur Pigou (1877–1959), assumed that welfare could be measured in terms of the utilities assigned by the economic agents to allocations of available resources in the economy. Because they treated utilities as cardinal measures, and therefore addable, the previous equation then becomes

$$W_x = u_1(x) + u_2(x) + \ldots + u_n(x),$$

where u_i stands for the utility assigned to the allocation x by the representative individual i. To use a familiar metaphor, let x_1 stand for a particular division of the pie that we want to compare with other ways of cutting it, say, x_2. If $W_{x_2} > W_{x_1}$ then x_2 should be preferred to x_1. Thus, the hypothetical problem mentioned above now becomes the practical one of finding the particular allocation of resources (let us call it x^*) that maximizes this fully specified function. The philosophical background of this approach is clearly utilitarian. Although Bentham was aware of the methodological difficulties attendant on summing up disparate utilities, the Pigovian welfare function evokes Bentham's maxim: "the greatest happiness for the greatest number of people."

This model did not remain dominant for very long. In light of criticisms leveled at it by Lionel Robbins in the 1930s, it became accepted that cardinal measures of utility cannot be objectively assessed. (This argument was later revised by positing that the only valid data are "revealed preferences," i.e., the observed choices made by individuals on the basis of inaccessible and noncomparable psychological dispositions.) The alternative is to use only ordinal ranking; the "New Welfare Economics" was established on this epistemological foundation. The touchstone of this approach is the concept of Pareto efficiency. An economic outcome *A* is more efficient in a Paretian sense than an outcome *B* if at least one agent is better off in *A* and no one is worse off; a Pareto optimum is a situation in which no one can be made better off without making at least one other agent worse off. Although informationally poor, this concept owes its success to the fact that it can be shown that the competitive market equilibrium happens to be a Pareto optimum. (This is known as the first "fundamental theorem of welfare economics.") There is, however, nothing intrinsically desirable, let alone just, about Paretian efficiency, except perhaps that it can be considered to be a rather robust protection against envy. To allow for transfers that could compensate the "losers" for the gains achieved by their more advantaged counterparts, several options were proposed (e.g., the Kaldor-Hicks compensation criteria). These suffer, however, from two drawbacks. One is of an ethical nature: The transfers envisioned by Nicholas Kaldor (1908–1986) or John Hicks (1904–1989) are purely hypothetical—in other words, welfare economics says nothing about the *actual* achievement of a fair redistribution. The second weakness is methodological; it can be shown that the criteria proposed by Kaldor or Hicks lead to intransitive results—transfers can lead to a vicious circle where one moves from situation *A* to situation *B* and then to situation *C,* only to come back to *A*.

A more promising development was proposed in 1938 by Abram Bergson, and later further developed by Paul Samuelson in the form of the Bergson-Samuelson social welfare function. *If* policymakers can assume that some particular end is socially desirable (there are, of course, many such ends, including Benthamite utilitarianism), then it is possible to build a social welfare function based only on ordinal rankings. At least in principle, such a function should be instrumental in identifying the optimal allocation of resources given the end chosen and given a "production–possibility frontier."

Impossibility Theorems or the Achilles's Heel of Welfare Economics?

In the early 1950s, Kenneth Arrow proposed a theorem that dashed the hope of ever finding an appropriate social welfare function. Most people would agree that in a democracy the electoral process is the only way to identify the criteria defining such a function. What Arrow demonstrated was that there is no paradox-free method of aggregating individual preferences into a social choice that expresses a collective preference for some particular way of allocating societal resources. If one assumes that any democratic procedure must meet a few minimum conditions, such as (a) "unlimited domain"—that is, all alternatives are allowed; (b) Pareto optimality—in other words, if every individual prefers *a* to *b,* then the social choice is also *a;* and (c) "independence of irrelevant alternatives"—that is, if *a* is preferred to *b* when there is a choice between *a, b,* and *c, a* should still be preferred to *b* when a fourth alternative, *d,* is added to the mix; thus, when there are more than two agents who must choose among three or more alternatives, only a "dictator" (i.e., someone whose

preferences prevail over those of everyone else) can break potential intransitive cycles (see above). But, evidently, the presence of a "dictator" is inadmissible in a democratic polity. Soon after Arrow proposed his theorem, several other comparable impossibility theorems were formulated.

The simplest response is to relax some of Arrow's initial conditions. For example, it can be shown that when all the alternatives can be aligned on a single axis (e.g., a left-to-right ideological spectrum), then cycling is no longer an issue. Behavioral welfare economics is another response. It allows for "irrational" preferences when such preferences can be shown to be empirically observed. But more radical, albeit more controversial, revisions of the research program of welfare economics are now under way.

Beyond Welfarism

A "welfarist" perspective assumes that the ranking of preferences over individual utilities are the building blocks of any welfare economics model. In recent years, several theorists have sought to move beyond this limitation by taking into account not only a wide range of dimensions of well-being, such as health or education, but also normative considerations having to do with the exercise of individual rights, such as the capacity to freely choose how one wishes to live. When such sources of information become available, interpersonal comparisons are no longer ruled out. The point is not that such measures are not themselves also partly subjective but rather that, contrary to utilities, they are not incommensurable —that is, intersubjective agreement on the dimensions of well-being is feasible. Reasoned debates can produce consensual agreement about states of the world that are more or less desirable (e.g., famine is to be avoided); democratic procedures other than voting open up, and Arrow's disturbing conclusion no longer appears so formidable. Many economists are still hesitant to move in that direction, but as the saying goes, "The train has already left the station." Nonwelfarism is likely to become the framework within which the future of welfare economics will be decided. Happiness economics encompasses various empirical approaches to the measurement and analysis of well-being. A more theoretically sophisticated path has been opened up by Amartya Sen's "capabilities approach." Sen distinguishes between "functionings" and "capabilities": The former enter into the definition of actual well-being, while the latter are a vector of functionings that a particular individual under specific circumstances has access to. For Sen, the goal of social policy is to equalize capabilities by making it possible for all to aspire to the functionings that they ought to be entitled to achieve as human beings deserving decent lives.

New Dilemmas

Nonwelfarism, however, raises some serious questions. While utilitarianism has no place for rights, nonwelfarist approaches risk falling into the trap of paternalism by arbitrarily selecting which rights or values, including the priority granted to freedom and self-realization, ought to be pursued. Moreover, the implementation of such schemes could challenge the political legitimacy and exceed the fiscal capacity and administrative resources of governments.

Laurent Dobuzinskis

See also Capabilities; Philosophy of Economics, History of; Pareto Optimality; Preference; Rational Expectations; Sen's Paretian Liberal; Social Choice Theory

Further Readings

Arrow, K. (1963). *Social choice and individual values* (2nd ed.). New York, NY: Wiley.

Atkinson, A. B. (2001). The strange disappearance of welfare economics. *Kyklos, 54,* 193–206.

Bergson, A. (1938). A reformulation of certain aspects of welfare economics. *Quarterly Journal of Economics, 52*(2), 310–334.

Frey, B. S., & Stutzer, A. (2002). *Happiness and economics.* Princeton, NJ: Princeton University Press.

Ng, Y.-K. (2004). *Welfare economics: Toward a more complete analysis.* New York, NY: Palgrave.

Pigou, A. C. (1932). *The economics of welfare* (4th ed.). London, England: Macmillan.

Robbins, L. (1962). *An essay on the nature and significance of economic science* (Rev. 2nd ed.). London, England: Macmillan.

Sen, A. K. (1999). *Commodities and capabilities.* Oxford, England: Oxford University Press.

Sen, A. K. (2009). *The idea of justice.* Cambridge, MA: Belknap Press.

We-Mode, Tuomela's Theory of

The *we-mode* theory originated in Raimo Tuomela's work and provides a philosophical and conceptual account of human social action and institutional reality. The theory accounts for phenomena such as collective activity, cooperation, and coordination in group contexts, as well as the mode of existence of institutional and corporate entities. The we-mode theory puts forth an ontologically individualistic (or *interrelationistic*) but conceptually collectivistic theory of *collective intentionality* that accounts for the irreducibility and apparent purposiveness of joint action and other collective intentional phenomena, such as collectively accepted institutional beliefs and collective emotions. The we-mode theory also relates to issues in social ontology concerning the mode of existence of social entities by providing a functional account of groups as social systems, abstractly formulated task-right systems with impersonal roles or positions that may be filled by any qualified members. The we-mode theory is accordingly conceptually equipped to account for the unity of institutions and organizations such as trade unions and business corporations, as well as institutional practices such as economic exchange and agreement making. The paradigms of we-mode thinking and acting are nevertheless to be found in simple joint actions where the participants intentionally cooperate as a group, as when singing a duet or painting a house together. This entry discusses the main conceptual features of Tuomela's we-mode theoretical framework and contrasts it with the *progroup I-mode,* ending with a brief discussion of recent developments within the we-mode theory and prospects for further research.

The We-Mode Theory

The we-mode theory may be seen as providing a revisionary conceptual framework of agency, which complements the more individualistic frameworks that are currently prominent in social science. The we-mode theory approaches social reality from an intuitive macrosocial or top-down perspective, where groups are considered the primary agents and individual members are viewed as position holders within those groups rather than as autonomous individuals who reason from their private points of view. The members of we-mode groups are assumed to give up part of their natural authority to act to the group and to be collectively committed to the group's activities. They are naturally disposed to perform their duties as group members and to conform to the directives, which are conferred on them by their collective decisions or by their leaders, depending on whether we are dealing with democratically or hierarchically organized groups, respectively. The most central authoritative element within the group is what Tuomela calls its *ethos,* which consists of constitutive contents such as goals and beliefs that have been collectively accepted by the members *for the group*. The collective construction of the ethos grants to groups a kind of intentional autonomy over and above its members, since collective acceptance is based on implicit or explicit agreement making, including compromises and interaction effects that cannot be reduced to the members' intentional psychologies. The group is irreducible to its members also for the further reason that it may survive partial or complete change of members, provided that the new members accept membership and are prepared to go along with the group's activities.

The we-mode framework is in itself ambivalent between an instrumentalist and a realist interpretation of group agency, but there are reasons to take it as a genuine explanatory theory rather than as a mere conceptual framework for intentional interpretation. The central motivating idea behind the we-mode framework is that groups often appear to behave in a purposive and goal-directed manner, thus warranting their treatment as agents at least in an instrumental sense. The we-mode theory is nevertheless in the last instance vindicated by its consequences rather than by its intuitive plausibility, and the onus is accordingly on showing that the behavioral predictions and explanations that are generated within the we-mode approach differ from the ones that are generated within the I-mode. The theory must of course grant that groups are not the same kinds of agents as individual agents, since they lack the corporeal and qualitative characteristics of ordinary human individuals. They may still be treated functionally as agents, which are real and active causal forces in the social world. This is all consistent with individuals or (in the last instance) their neural states and behavioral dispositions being the fundamental causal motors in the social world.

Thus, we may say that the intentional agent in the we-mode approach is the group but the ontological agent is the individual.

The we-mode framework is based on an intuitive analogy between individual and collective agents, which may be elucidated in terms of the three criterial features of the we-mode: *group reasons, collective commitment,* and the *collectivity condition.* These constitutive features of the we-mode framework are argued to be interdependent but irreducible to one another and to the standard individualist framework of agency. The we-mode framework is thus irreducible to the individualist (I-mode) framework of agency in a manner that is not entirely unlike the way in which the individualist framework of intentional interpretation has been argued to be irreducible to the vocabulary of the physical and biological sciences. The difference between the two cases is that in the latter we are dealing with reduction between two different conceptual schemes (viz., the conceptual framework of the physical and biological sciences against the conceptual framework of intentional interpretation), whereas in the former we are dealing with reduction between two different applications of essentially one and the same conceptual scheme (viz., groups vs. individuals as intentional agents).

The analogy view regards groups as agents that may act instrumentally in light of the motivating reasons that are provided by the collectively accepted beliefs, desires, and intentions that are contained in its ethos. These reasons are normally thought to provide the members with reason-based directives, which are *preemptive* (viz., exclusionary with regard to competing individualistic I-mode reasons) and *presumptive* (viz., sufficient in the absence of stronger countervailing reasons to the contrary). The notion of group reasons may be further elucidated by drawing a distinction between (a) the *group agent's reasons* relative to its ethos and (b) the group-based reasons (or *group reasons*) of its members. The group agent's reasons are provided by the collectively accepted contents in its ethos, whereas the group reasons of its members are the proximate reasons that they act for when they are performing their duties as group members. The distinction is especially relevant in the case of hierarchically organized groups, where only a proper subset of the members have insight into the group's constitutive goals but other members are prepared to go along with the group's activities and follow the directives that are conferred on them by the leaders. The group members' reasons may also be seen to involve elements that are not involved on the level of the group agent's reasons, because the members are normatively and group-socially obligated to one another to perform their parts as group members. The we-mode theory accordingly posits a kind of normative emergence when we move from the group agent's reasons to the group reasons of its members, according to the most recent formulations of the theory by Tuomela.

The satisfaction of an intention requires commitment to bringing about the content of the intention, and the group members may thus be required to be collectively committed to their joint intention, according to the analogy view. The notion of commitment entails a disposition to think and act in an appropriate manner in order to bring about or maintain the content of the intention, namely, to see to it that the content of the intention is realized. This may require performing relevant practical inferences in addition to performing the right bodily actions, and collective commitment in particular requires the agents to take the group into account in their practical reasoning and deliberation (e.g., "We will perform *X*"; "Our performing *X* requires that I do *Y*"; "Hence, I will do *Y*"). The members of a we-mode group are group-socially committed to one another to perform their parts in a manner that incurs properly normative rights and obligations between them, as remarked in the discussion of group reasons above. The members may be expected to step in and help one another if needed and to perform whatever is required for the satisfaction of the collective goal, even though that may require performing more than their distributive share (in the limiting case, the entire collective action may be performed by a single member, while the rest of the group monitors the situation and is ready to step in if needed). This results in more solidarity and stability being present in the we-mode collective agency than in the I-mode, where the members are required to deliberate separately when the division of tasks changes.

The satisfaction conditions for the group's intention are fixed by the content of the intention and by the *collectivity condition,* according to which the collective intention is satisfied simultaneously and interdependently for all individual members of the group on conceptually necessary grounds. Tuomela, in his

Philosophy of Sociality, allows there to be weaker *progroup I-mode* counterparts to the central notions of collective commitment, group reasons, and the collectivity condition, where the modal component does not hold true. Thus, a collection of individuals may each separately intend to realize the same outcome by means of their coordinated behaviors, but here the simultaneous satisfaction of their intentions is a contingent and situation-specific matter. They do not intend to act together as an irreducible social whole, that is, as a group or as a body (metaphorically speaking), and therefore they are not bound together in the strong we-mode manner. There thus exist both we-mode and progroup I-mode collective intentionality, and both are required for providing a complete account of social life. The we-mode may be seen to provide the fundamental common ground of human society, but even we-mode social institutions tolerate an important amount of activity in the I-mode: One need only think about what would happen to social institutions such as the economy if people suddenly stopped using money (a we-mode activity or constitutive institutional rule), although the economic sphere is usually thought to consist primarily of egocentric activity.

The We-Mode and the Progroup I-Mode

The main conceptual difference between the we-mode and the (progroup) I-mode is that in the we-mode the conceptual and justificatory direction is top down—from the group level to the member level—whereas in the I-mode, the conceptual and justificatory direction is bottom up—from the member level to the group level. The macrosocial perspective that the we-mode framework incorporates may accordingly be unpacked in terms of a distinction between three levels of explanation and the asymmetrical relations of reduction or explication between them: (1) the macro level of social institutions and collective social reality, including group agents' intentions and their reasons for them; (2) the interpersonal meso level of joint intentions and actions, including group members' collective commitment to their joint actions and group-social normativity between them; and (3) the personal level of (I-mode or we-mode) mental states and attitudes, including the members' part-performance intentions and their coordinative social reasons for performing their shares. The we-mode accounts for (3) in terms of (2) and (1), whereas the I-mode accounts for (1) and (2) in terms of (3). The central opposition in the I-mode/we-mode theory is accordingly not between individual and collective agency but between *private* and *collective* agency—that is, agency that is based (in a conceptual sense) on an individual agent's private mental states and attitudes against agency that is based in part on the group's actions and attitudes. Thus, both I-mode and we-mode mental states are *personal* states in the sense that they are attributed to individuals, but we-mode states are essentially public and group dependent, whereas I-mode states are private and atomistic.

The conceptual relations between the group level, the interpersonal jointness level, and the individual agent level are accounted for in partly circular terms when we are operating within the we-mode framework. The central member-level counterpart of the group's intention is the personal *we-intention,* which may upon analysis be seen to presuppose the existence of a joint intention between the members. Technically, a we-intention may in the present context be understood as a *reason-based we-attitude,* where the others' thinking and acting thus and so is a partial motivating or presuppositional reason for one's own thinking and acting thus and so. Thus, when the members we-intend to perform a joint action (*X*) together, each member intends to perform his part of *X as his part of X,* in part because of the social reason that the others intend to perform their parts of *X*. The circularity of the account results from the fact that the *X* term seemingly cannot be analyzed without reference to the joint intention itself should the entire joint action and not only its parts be intentionally performed. This kind of circularity is nevertheless functionally innocuous because the members may operate on a cognitively nondemanding and pre-analytic notion of joint intention. The group intention, the joint intention, and the we-intentions of the participants taken together may be viewed as equivalent under ideal conditions when each participant is functioning as a full group member; by relaxing this condition, we arrive at quasi we-mode groups, where some of the members function in the I-mode. The modest circularity of we-intentions is not only tolerable but may even be regarded as a vindication of the conceptual irreducibility of the we-mode framework, since it is a characteristic of irreducible concepts that they cannot be accounted for without circularity arising.

The collective attitudes of the group are accounted for within the we-mode approach in terms of the members' reflexive and performative *collective acceptance* of attitudes for the group. Collective acceptance is *performative* in the sense that it creates attitudes for the group on an analogy with declarative speech acts, which have the power to create institutional facts by representing them as true (e.g., "I hereby pronounce you husband and wife"). Collective acceptance is *reflexive* in the sense that it involves reference to itself within its representational content (e.g., "We hereby collectively accept that squirrel pelt counts as money in virtue of our collectively accepting that squirrel pelt counts as money in the group"). The conceptual irreducibility of the we-mode may accordingly be partially brought out in terms of the need for the group members to have a thick reflexive concept of "we" that is operative in the collective construction of attitudes for the group. This strong "togetherness" notion of "we" may be pulled apart by philosophical analysis from the weak aggregative notion of "we" that I-mode social groups are based on, although ordinary language does not make such a distinction. The difference between the two is crucial for understanding human sociality, if the mainly functional arguments that have been put forth in support of the I-mode/we-mode distinction are tenable. The present entry has concentrated on the main conceptual features of the we-mode framework, and the reader is referred to the original works by Tuomela for a closer survey of his arguments in favor of the we-mode.

Recent Developments in the We-Mode Theory

There have been two important foci of research in recent years within the we-mode approach: (1) group-internal normativity and the authoritative nature of group reasons and (2) the relation between the we-mode framework and team reasoning in game theory. The following discussion concentrates on the relation between team reasoning and the we-mode approach, since recent research on group-internal normativity was already discussed above in the context of group reasons and group-social commitment. Within the I-mode/we-mode theory, individuals who function in the I-mode think and act on the basis of their own beliefs and preferences, whereas in the we-mode, they think and act in terms of their collectively accepted attitudes and may therefore be characterized as being engaged in *we-reasoning*. To elaborate, whereas in the I-mode the primary practical question is "What should I do?", in the we-mode the primary practical questions are "What should our group do?" and "What should I do as a group member?" As has been noted by economists and game theorists, framing a decision problem in terms of *team thinking* may help agents achieve collectively rational outcomes in coordination problems such as the Hi-Lo and in social dilemma situations such as the Prisoner's Dilemma. Relating their notion of we-reasoning to the game-theoretic notion of team reasoning, the following claims have been put forth by Hakli, Miller, and Tuomela: (a) operating in the we-mode may lead to different action recommendations from the progroup I-mode, and it may prevent dilemma situations from arising because the agents frame the situation as a collective-decision problem; (b) we-reasoning (in the we-mode) can decrease the amount of Nash equilibria in decision problems, but it cannot increase them, and it may thereby increase the likelihood of collectively rational solutions and lead to greater collective order and stability than (progroup) I-mode reasoning; and (c) we-reasoning (in the we-mode) economizes the members' cognitive activities since they may proceed on the assumption that the others will perform their parts and are thus not required to be aware of the others' specific mental states to the same extent as in the (progroup) I-mode. Nevertheless, there remain substantial problems about relating the impoverished conceptual framework of game theory to the richer action-theoretic concepts that are employed in philosophy of action and within the we-mode approach.

Matti Heinonen

See also Collective Agents; Collective Emotions; Collective Intentionality; Collective Moral Responsibility; Collective Rationality; Group Beliefs; Social Institutions; Social Ontology, Recent Theories of; Speech Acts; Team Reasoning

Further Readings

Bacharach, M. (2006). *Beyond individual choice: Teams and frames in game theory* (N. Gold & R. Sugden, Eds.). Princeton, NJ: Princeton University Press.

Bratman, M. (1993). Shared intention. *Ethics, 104,* 97–113.

Hakli, R., Miller, K., & Tuomela, R. (2010). Two kinds of we-reasoning. *Economics & Philosophy, 26,* 291–320.
Searle, J. (1990). Collective intentions and actions. In P. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 401–416). Cambridge: MIT Press.
Searle, J. (2010). *Making the social world: The structure of human civilization.* Oxford, England: Oxford University Press.
Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behavior. *Social Philosophy & Policy, 10,* 69–89.
Tuomela, R. (2005). We-intentions revisited. *Philosophical Studies, 125,* 327–369.
Tuomela, R. (2007). *The philosophy of sociality.* New York, NY: Oxford University Press.
Tuomela, R. (in press). Group reasons. *Philosophical Issues.*
Tuomela, R., & Miller, K. (1988). We-intentions. *Philosophical Studies, 53,* 115–37.

World-Systems Analysis

This entry introduces the influential social-scientific systemic approach known as *world-systems analysis* and reviews its methodological precepts, showing their philosophical import.

World-systems analysis emerged from the context of Dependency Theory, in response to modernization theory, which suggested that failure to develop was the fault of nations themselves, not of systemic forces. Modernization theory argued that development would succeed if the less developed simply acquiesced in economic and political programs favored by and allegedly followed in the West. Dependency scholars suggested that development and underdevelopment were two sides of the same coin, acting to develop a core and underdevelop a periphery. When critical scholars turned to broaden their study of core and peripheral social structures and trace their historical origins and relations, world-systems analysis emerged.

Philosophical and Methodological Precepts

World-systems *analysis* is not a theory but an approach that includes several philosophical and methodological positions, being (a) systemic (world economies, not nation-states and/or firms, are the appropriate unit of analysis), (b) materialist (focused on economic processes, especially capital accumulation), (c) transdisciplinary (emphasizing holistic coherence by transcending traditional academic disciplines), (d) historical (deploying historical method as intrinsic and extending research back 500 years or more), and (e) methodologically open (and nondeterministic).

Let us view these methodological positions in some detail:

a. *System:* World-systems analysis escapes the limitations of state centrism. During the Romantic Period, the history of most Western nations was recast into a statecentric mold. History was rewritten to portray states as self-contained units of development. Influenced by the "total-history" approach and the perspective of the *longue durée* pioneered by Fernand Braudel in the 1940s, world-systems analysis rejected the state as the primary unit of analysis, in favor of new concepts. Dependency theory and world-systems analysis agree that the whole is greater than the sum of its parts and that no part (e.g., nation-state) can be adequately understood in isolation from its position, role, and function in the whole—in other words, the world-system and its historical structure and processes.

b. *Economic materialism:* A materialist perspective informs world-systems analysis by a focus on capital accumulation as a force animating individuals, families, firms, states, and systems. Material wealth derived from the production of commodities for exchange on the world market was redeployed to reshape social structures and ensure future accumulation. World-systems analysis contends that the origins of capitalism emerged in the European world economy in the "long 16th century" (1450–1600). This period of "transition to capitalism" is understood as a modal shift from a system where territorial power yielded economic wealth to one where capitalist wealth yielded (geo)political power. Once consolidated, this new "logic" constituted the central dynamic of the "modern world-system," understood to be distinct from that of any earlier period. In a break with this interpretation, a group of scholars led by Andre Gunder Frank and Barry K. Gills challenged the idea that the critical structuring element of the world system emerged with the transition to capitalism within Europe, and highlighted

continuities in structure and process before and after the alleged transition to capitalism. This continuity included nearly all the critical elements attributed to the modern capitalist world-system. A major disagreement emerged. The new world system approach (note the lack of the hyphen) emphasizes the thesis that the world system, and its main structure and processes, existed well before 1500 and the rise of European hegemony. This perspective attracted a multidisciplinary circle of scholars in prehistory, world history, anthropology, sociology, geography, political science, and international relations.

c. *Holism/transdisciplinarity:* Both of the world(-)systems approaches outlined above accept the proposition that holistic analysis of social phenomena is preferred. Development cannot be adequately apprehended without full attention to history, politics, sociology, or anthropology. As part of world(-)systems analysis, scholars considered the separation of the social sciences and the consequent compartmentalization and fragmentation of knowledge, which, however, has generated the impetus to reintegrate insights into a new coherent structure.

d. *History:* World(-)systems analysis is a historical method that attempts to trace the origins of capitalism and capture its dynamics in world history. To do so, some advocate a profound differentiation between precapitalist and capitalist social systems, and for some the "long 16th century" remains the pivotal transitional period in world history. The chronology and periodization of the critical junctures of the world(-)system have never been agreed on. Some scholars suggest that capitalist practices were in place by the 14th or 15th century, while others push similar dynamics back into the 12th and 13th centuries and eastward into the Orient. A more radical interpretation pushes the advent of capitalist processes back to the classical period, or even deep into prehistory. Another tendency of world(-)systems analysis adopts the comparative method to look at social systems, including the study of very different forms and scales of world-systems and world economies, especially in microsystems and hinterlands.

e. *Methodological openness:* World(-)systems analysis adopts a nondeterministic mode, one that emphasizes cyclical conceptions of time and historical processes while not privileging linear conceptions of social change. Not all processes recur, but world(-)systems analysis sees great importance in the repetition of systemic cycles and patterns, such as economic and political rise and decline, the concentration and de-concentration of wealth and power, and cycles or transitions of hegemony. Chaotic processes may give rise to various cycles and trends, but they remain indeterminate in the long term. There is little agreement, however, on what might come next. Following Karl Marx, some argue that the contradictions inherent in the capitalist world economy will eventually lead to its final breakdown and transformation, perhaps into a form of socialism, but the essential elements of any subsequent social systems cannot be predicted. Other predictions suggest a long-term shift in the center of gravity of the world system back toward Asia, and China in particular. There is an emphasis on the utility of world(-)systems analysis to understanding the contemporary global crisis as a global center shift and a hegemonic transition, as well as a civilizational crisis.

Barry K. Gills and
Robert A. Denemark

See also *Annales* School; Capitalism; Historicism; Norbert Elias: Process of Civilization and Theory of Sciences; Philosophy of History; Systems Theory

Further Readings

Denemark, R., Friedman, J., Gills, B., & Modelski, G. (Eds.). (2000). *World system history: The social science of long-term change*. London, England: Routledge.

Frank, A. G. (1998). *ReOrient: Global economy in the Asian age*. Berkeley: University of California Press.

Gills, B. (Ed.). (2010). *Globalization in crisis*. London, England: Routledge.

Wallerstein, I. (2011). *The modern world-system: Vol. 1. Capitalist agriculture and the origins of the European world-economy in the sixteenth century, with a new prologue*. Berkeley: University of California Press.

Index

Entry titles and their page numbers are in **bold.** Page numbers preceded by 1: are in volume 1, and page numbers preceded by 2: are in volume 2.