FOURTH EDITION

PHILOSOPHY OF SOCIAL SCIENCE

ALEXANDER ROSENBERG

Philosophy of Social Science

FOURTH EDITION

Philosophy of Social Science

Alexander Rosenberg
Duke University

WESTVIEW
PRESS
A Member of the Perseus Books Group

Westview Press was founded in 1975 in Boulder, Colorado, by notable publisher and intellectual Fred Praeger. Westview Press continues to publish scholarly titles and high-quality undergraduate- and graduate-level textbooks in core social science disciplines. With books developed, written, and edited with the needs of serious nonfiction readers, professors, and students in mind, Westview Press honors its long history of publishing books that matter.

Published by Westview Press,
A Member of the Perseus Books Group

Find us on the World Wide Web at www.westviewpress.com.
Every effort has been made to secure required permissions for all text, images, maps, and other art reprinted in this volume.

Library of Congress Cataloging-in-Publication Data
Rosenberg, Alexander, 1946–
Philosophy of social science / Alexander Rosenberg.—4th ed.
p. cm.
ISBN 978-0-8133-4592-5 (pbk. : alk. paper)—
ISBN 978-0-8133-4593-2 (e-book)
1. Social sciences—Philosophy. I. Title.
H61.R668 2012
300.1—dc23
2011041428

10 9 8 7 6 5 4 3 2 1

To the memory of Blanca N. Rosenberg,
mentor and teacher to two generations of students
in the School of Social Work, Columbia University, 1958–1979

Contents

Preface to the Fourth Edition ix

1 WHAT IS THE PHILOSOPHY OF SOCIAL SCIENCE? 1

2 THE METHODOLOGICAL DIVIDE: NATURALISM VERSUS INTERPRETATION 11

3 THE EXPLANATION OF HUMAN ACTION 35

4 INTENTIONALITY AND INTENSIONALITY 55

5 BEHAVIORISM IN THE BEHAVIORAL SCIENCES 71

6 PROBLEMS OF RATIONAL CHOICE THEORY 89

7 SOCIAL PSYCHOLOGY AND THE CONSTRUCTION OF SOCIETY 117

8 PHILOSOPHICAL ANTHROPOLOGY 139

9 HOLISM AND ANTIREDUCTIONISM IN SOCIOLOGY AND PSYCHOLOGY 169

10 FUNCTIONALISM AS A RESEARCH PROGRAM 191

11 SOCIOBIOLOGY OR THE STANDARD SOCIAL SCIENCE MODEL? 207

12 THEORIES OF CULTURAL EVOLUTION 233

13 RESEARCH ETHICS IN SOCIAL INQUIRY 253

14 FACTS AND VALUES IN THE HUMAN SCIENCES 273

15 SOCIAL SCIENCE AND THE ENDURING QUESTIONS OF PHILOSOPHY 293

Bibliography 305
Index 311

Preface to the Fourth Edition

A generation is a lifetime in the history of a textbook. I wrote the first edition of this work in the late 1980s, long before the advent of online resources and well before the handbooks, encyclopedias, companions, and guides to every discipline and subdiscipline began to proliferate. The persistence of an old-fashioned conventional introduction to the philosophy of social science in the online informational environment of the twenty-first century requires explanation. Some hypotheses are disobliging and disconfirmed: it can't be that no one cares enough about the subject; there are, after all, by Google Scholar's account, eleven books with titles that are variations on *The Philosophy of Social Science.* Nor can it be that no one plugged into the Web is interested in this subject or my take on it: I once had to threaten legal action to get a plagiarized version of the second edition taken down from a website.

To explain the persistence of this book across twenty-four years and four editions I propose that we consider this hypothesis: its particular approach to the philosophy of social science has persisted because of the merit of the book's central premises: that social scientists must take sides on philosophical problems, whether they like it or not, even whether they know it or not, and that the problems of the philosophy of social science are all versions of one or another fundamental problem of philosophy—problems of metaphysics, of epistemology, of ethics.

A great deal has changed in the social sciences since the eighties, when the first edition was gestating. Economics, for example, has been greatly changed, because of both events in the economy and changes in social sciences it long refused to take seriously, especially in cognitive social psychology, experimental economics, and evolutionary game theory. Economics has also succumbed to attacks on its moral neutrality and indifference to application in human development. These too have made that field recognizably more like the other social sciences it used to pretend to distain.

Meanwhile, anthropology, sociology, and social psychology, along with politics, have been swept up in a tsunami of Darwinian analysis originating from tectonic changes in evolutionary biology and biological anthropology. Owing to the zeal—perhaps *trop de zèle*—of the second and third generation of sociologists, evolutionary psychologists, and gene-culture coevolutionary theorists, there is after thirty or forty years still no end in sight to the Darwinization of the social sciences.

Another of the great changes in social science since the first edition of this book is the increasing willingness of European students of human affairs to be influenced by naturalistic, empirical, and data-based approaches to social science. The empiricist and quantitative approaches to the sciences of human affairs had their origins in European thinkers—Durkheim, Weber, Walras. But that approach was eclipsed and effaced in the middle years of the twentieth century by Marxism, the Frankfurt School, phenomenology, structuralism, postmodernism, deconstruction, and critical theory. Now the movement that is here called philosophical anthropology is moving back toward an appreciation of the older European tradition of social research carried through the twentieth century by English-language social science. But the Europeans bring their intellectual tradition along to this new exchange. This raises questions of compatibility that few address.

All the social sciences have become much more sensitive to, and much more influenced by, theories and findings that reflect the experience and perhaps also the special information—if not exclusive knowledge—of women, ethnic, racial, and other hitherto and still largely marginalized groups. How to incorporate these new voices and thoughts remains a vexed matter.

However, these changes in the social sciences have brought along with them not so much changes in the philosophical questions they raise, but a new vocabulary with which to express the persistent philosophical questions that face the social scientist. This edition reflects the new vocabulary of the human sciences, while continuing the previous editions' insistence that the problems social science faces are old wine in new bottles, but just as intellectually intoxicating as ever.

Readers of previous editions will find much new material on the role of models and equilibrium explanations in economics; new discussions of how speech acts create norms and thus construct social practices and institutions; treatment of the work of Foucault, Bourdieu, and Habermas as continuing a tradition that began with Hegel; the problem of spontaneous order in the creation of institutions; and the relationships of Rawls's moral theory to social research and Sen's capacity theory to the broad problem of how facts and values intersect.

Veteran readers will, I hope, find the prose generally clearer, since the writing of each edition reflects less and less patience with longer and longer sentences.

As in previous editions, I begin with an explanation of why philosophy is relevant to the human sciences, and then I explore the problems raised by alternative explanatory strategies of the human sciences. In the twentieth century these problems spawned familiar theoretical and methodological movements: behaviorism, structuralism, and a variety of interpretational theories including critical theory, to name only a few. Even among social scientists who accepted no labels for their views, these problems facing their explanations led to significant shifts in the aims and methods advocated for the social sciences. Despite the changes described briefly above, the challenges facing social science in the twenty-first century remain the same as those that confronted sociologists such as Emile Durkheim and Claude Lévi-Strauss, psychologists such as B. F. Skinner, economists such as Milton Friedman, and cultural theorists such as Michel Foucault or Pierre Bourdieu. The thinking of all these figures and others is sketched in this book, where it confronts fundamental problems of method and theory raised by the philosophy of science.

The major departure from previous editions of this work is to be found in its organization. Previous editions structured this introduction as a sustained argument in eight chapters. Although this had its dialectical advantages, it came with pedagogical drawbacks. Having conceded priority to its role as a textbook, I have reorganized this edition into fifteen chapters whose titles make very clear exactly how they relate to the agenda of problems treated in a one-semester academic course.

Previous editions mention my debts to many scholars—social scientists and philosophers. The ones whose lessons have stuck with me the longest include David Braybrooke, Donald T. Campbell, Martin Hollis, Jonathan Bennett, Dan Hausman, Harold Kincaid, Martin Trow, Alasdair McIntyre, and Amartya Sen.

CHAPTER ONE

What Is the Philosophy of Social Science?

Most sociologists and anthropologists will agree on the definition and the domain of their disciplines; the same holds true for many psychologists, political scientists, and almost all economists. The same cannot be said for philosophers and philosophy. Philosophy is a difficult subject to define, which makes it difficult to show social scientists why they should care about it—the philosophy of social science in particular. This chapter provides a definition of philosophy that makes the subject inescapable for the social scientist. It shows that, whether as an economist or an anthropologist, one has to take sides on philosophical questions. One cannot pursue the agenda of research in any of the social sciences without taking sides on philosophical issues, without committing oneself to answers to philosophical questions. At a minimum, social scientists need to recognize this fact about themselves. It is even better if the choices made are based on evidence and argument.

WHAT IS PHILOSOPHY?

Philosophers do not agree among themselves on the definition of their subject. Its major components are easy to list, and the subjects of some of them are even relatively easy to understand. The trouble is trying to figure out what they have to do with one another, why combined they constitute one discipline—philosophy—instead of being parts of other subjects, or their own independent areas of inquiry.

The major subdisciplines of philosophy include logic, the search for well-justified rules of reasoning; ethics (and political philosophy), which concerns itself with right and wrong, good and bad, justice and injustice, in the conduct of individuals and states; epistemology or the theory of knowledge, the inquiry into the nature, extent, and justification of human knowledge; and metaphysics, which seeks to determine the most fundamental kinds of things that exist in reality and what the relations between them are. Despite its abstract definition, many of the questions of metaphysics are well known to almost all people. For example, Is there a God? or, Is the mind just the brain, or something altogether nonphysical? or, Do I have free will? are all metaphysical questions most people have asked themselves.

But these four domains of inquiry don't seem to have much to do with one another. Each seems to have at least as much to do with another subject altogether. Why isn't logic part of mathematics, or epistemology a compartment of psychology? Shouldn't political philosophy go along with political science, and isn't ethics a matter ultimately for people who deliver sermons? Whether we have free will or the mind is the brain is surely a matter for neuroscience. Perhaps God's existence is something to be decided upon not by an academic inquiry but by personal faith. The question thus remains: What makes them all parts of a single discipline, philosophy?

The answer to this question organizes this book, and it is pretty clear. Philosophy deals with two sets of questions: first, questions that the sciences—physical, biological, social, behavioral—cannot answer now and perhaps may never be able to answer; second, questions about why the natural and social sciences cannot answer the first set of questions.

There is a powerful argument for this definition of philosophy in terms of its historical relationship with science. The history of science from the ancient Greeks to the present is that of one compartment of philosophy after another breaking away from philosophy and emerging as a separate discipline. But each of these separated disciplines has left philosophy with a set of distinctive problems, issues the discipline cannot resolve, but must leave either permanently or at least temporarily for philosophy to deal with: Mathematics leaves philosophy questions like, What is a number? Physics leaves to philosophy questions like, What is time? There are other questions science appears to be unable to address—the fundamental questions of value, good and bad, rights and duties, justice and injustice—that ethics and political philosophy address. Questions about what ought to be the case, what we should do, and what is right and wrong, just and unjust are called *normative*. By contrast, questions in science are presumably descriptive or, as is sometimes said, *positive*, not normative. Many of the normative questions have close cousins in the social and behavioral sciences. Thus, psychology will interest itself in why individuals hold some actions to be right and others wrong; anthropology will consider the sources of differences among cultures about what is good and bad; political science may study the consequences of various policies established in the name of justice; economics will consider how to maximize welfare, subject to the normative assumption that welfare is what we ought to maximize. But the sciences—social or natural—do not challenge or defend the normative views we may hold.

In addition to normative questions that the sciences cannot answer, there are questions about the claims of each of the sciences to provide knowledge, or about the limits of scientific knowledge, that the sciences themselves cannot address. These are among the distinctive questions of the philosophy of science, including questions about what counts as knowledge, explanation, evidence, or understanding. The philosophy of science is that subdiscipline of philosophy devoted to addressing these questions.

PHILOSOPHICAL PROBLEMS OF SOCIAL SCIENCE

If there are questions the sciences cannot answer and questions about why the sciences cannot answer them, why should a scientist, in particular a behavioral or social scientist, take any interest in them? The reason is simple. Though the sciences cannot answer philosophical questions, individual scientists have to take sides on the right answers to them. The positions scientists take on answers to philosophical questions determine the questions they consider answerable by science and choose to address, as well as the methods they employ to answer them. Sometimes scientists take sides consciously. More often, they take sides on philosophical questions by their very choice of question, and without realizing it. The philosophy of science may be able to vindicate those choices. At the least, it can reveal to scientists that they have made choices, that they have taken sides on philosophical issues. It is crucial for scientists to recognize this, not just because their philosophical positions must be consistent with the theoretical and observational findings of their sciences. Being clear about a discipline's philosophy is essential because at the research frontiers of the disciplines, it is the philosophy of science that guides inquiry.

As Chapter 2 argues, the unavoidability and importance of philosophical questions are even more significant for the social scientist than for the natural scientist. The natural sciences have a much larger body of well-established, successful answers to questions and well-established methods for answering them. As a result, many of the basic philosophical questions about the limits and the methods of the natural sciences have been set aside in favor of more immediate questions clearly within the limits of each of the natural sciences.

The social and behavioral sciences have not been so fortunate. Within these disciplines, there is no consensus on the questions that each of them is to address, or on the methods to be employed. This is true between disciplines and even within some of them. Varying schools and groups, movements and camps claim to have developed appropriate methods, identified significant questions, and provided convincing answers to them. But among social scientists, there is certainly nothing like the agreement on such claims that we find in any of the natural sciences. In the absence of agreement about theories and benchmark methods of inquiry among the social sciences, the only source of guidance for research must come from philosophical theories. Without a well-established theory to guide inquiry, every choice of research question and of method to tackle it is implicitly or explicitly a gamble with unknown odds. The choice the social scientist makes is a bet that the question chosen is answerable, that questions not chosen are either less important or unanswerable, that the means used to attack the question are appropriate, and that other methods are not.

Chapter 2 outlines the alternative choices, bets, and wagers about the best way to proceed that face the social scientist. When social scientists choose to employ methods as close as possible to those of natural science, they commit themselves to the position that the question before them is one that empirical science can answer. When they spurn such methods, they adopt the contrary view, that the question is different in some crucial way from those addressed in the physical or biological sciences. Neither of these choices has yet been vindicated by success that is conspicuous enough to make the choice anything less than a gamble.

Whether these gambles really pay off will not be known during the lifetimes of the social scientists who bet their careers on them. Yet the choices must be justified by a theory, either one that argues for the appropriateness of the methods of natural science to the question the social scientist addresses, or one that explains why these methods are not appropriate and supplies an alternative. Such theories are our only reasonable basis for choosing methods of inquiry in the social sciences.

But these theories are philosophical, regardless of whether the person who offers them is a philosopher or a social scientist. Indeed, social scientists are in at least as good a position as philosophers to provide theories that justify methods and delimit research. In the end, the philosophy of social science not only is inevitable and unavoidable for social scientists, but it must also be shaped by them as much as by philosophers.

The traditional questions of the philosophy of social science reflect the importance of the choice among these philosophical theories. And in this book we shall examine almost all of those questions at length. By contrast with this approach to social science, which very self-consciously takes its inspiration from the natural sciences, there are disciplines that make the meaning and intelligibility of human affairs central to their explanations. These social scientists (and the philosophers who embrace their aims and methods as the right way to proceed) contrast their commitment to understanding with demands for prediction. They are indifferent or hostile to the notion that their disciplines should provide predictive knowledge about individuals or groups. In Chapters 7 and 8 we look at this approach.

In Chapters 8 through 10 we also turn to questions about whether the primary explanatory factors in social science should be large groups of people such as social classes or communities and their properties—so-called structural properties, as Marx, Durkheim, and other social scientists have argued—or whether explanations must begin with the choices of individual, often "rational" human agents, as contemporary economists and some political scientists argue. The differences between the various social sciences, especially economics and sociology, on this point are so abstract and general that they have long concerned philosophers. The social scientist who holds that large-scale social facts explain individual conduct, instead of the reverse, makes strong metaphysical assumptions about the reality of groups independent of the individuals who compose them. Such a theory—called *holism*—also requires a form of explanation called *functionalism*, which raises other profound questions about differences between the explanatory strategies of social and natural science. As a theory that gives pride of explanatory place to social wholes, holism might seem quite unappealing. But the alternative to it, *individualism*, as advanced by economists, political scientists, and biologically inspired social scientists, for instance, also faces equally profound philosophical questions.

Problems of functionalism, holism, and individualism are exacerbated by the ever-increasing influence of biological science, and especially Darwin's theory of natural selection, on all the social and behavioral sciences. This is the subject of Chapters 11 and 12, which report on several lively debates at the intersection of biology and the social sciences and their philosophies.

In Chapters 13 and 14 we turn to the relation between the social sciences and moral philosophy. We examine whether we can expect the social sciences themselves to answer questions about what is right or fair or just or good. Many philosophers and social scientists have held that no conclusions about what ought to be the case can be inferred even from true theories about what is the case. Others have asserted the opposite. No matter who is right, it will still turn out that alternative approaches to social science and competing moral theories have natural affinities to, and make strong demands on, one another as well. We must also examine the question of whether there are morally imposed limits to legitimate inquiry in the social sciences.

Finally, in Chapter 15 we try to show why the immediate choices that social scientists make in conducting their inquiry commit them to taking sides on the most profound and perennial questions of philosophy. If this is right, then no social scientist can afford to ignore the philosophy of social science or any other compartment of philosophy.

ONE CENTRAL PROBLEM OF THE PHILOSOPHY OF SOCIAL SCIENCE

The central philosophical dispute about the scope, aim, and prospects for each of the social sciences taken separately, and all of them together, is what sort of knowledge they should or can seek. The debate takes place against a background argument about the nature of understanding in the natural sciences. There it is widely held that increases in understanding are certified by improvements in prediction.

Among social scientists who accept the requirement that their discipline provide the kind of knowledge natural science provides—demographic sociologists, econometricians, experimental social psychologists, or political scientists interested in voting behavior, for instance—there is a strong commitment to improving prediction as the mark of increasing understanding. Among social scientists there are debates about how reliable and precise their respective disciplines' predictions can be and whether they can get better. But other social scientists reject the demand that their discipline provide the same kind of understanding natural science offers. These social scientists offer alternative explanations of why their subjects cannot, and should not, seek predictive knowledge and improvements in it. They provide quite different accounts of what the aims and objectives of their disciplines can be.

The question centers on the fact that it is human beings, in groups and individually, whose behaviors, actions, and their consequences we are trying to understand that make the difference between natural and social science. It is what shapes the nature and scope of the knowledge social science can provide. Should the subject matter of these disciplines make the aims and methods of the social sciences as a whole radically different from those of the natural sciences?

The natural sciences are often alleged, especially by natural scientists and others impatient with social science, to have made far greater progress than the social sciences. Questions naturally arise as to why that is so and what can be done to accelerate the progress of social science toward achievements comparable to those of natural science. But one should notice that these two questions have controversial philosophical presuppositions: they presuppose (1) that we know what progress in natural science is and how to measure it; (2) that, based on our measurements, the natural sciences have made more progress; and (3) that the social sciences aim for the same kind of progress as the natural sciences.

If you agree that progress in the social sciences leaves much to be desired compared with the natural sciences, then you must be able to substantiate those three presuppositions. However, if you consider that the social sciences cannot or should not implement the methods of natural science in the study of human behavior, you will reject as misconceived the invidious comparison between the natural and the social sciences, along with the presuppositions on which it is based. But if you conclude that the study of human action proceeds in a different way and is appraised with different standards than the natural sciences, then you will have equally strong presuppositions about the aims and achievements of social science to substantiate.

Chapters 2 through 4 of this volume outline the arguments both for and against the claims that the social sciences have failed to progress and that this failure needs explanation. Both arguments have one view in common: a neat compromise is impossible. Such a compromise would suggest not that social science has made as much progress as have the natural sciences, but that it has made some. It would suggest that very broadly the methods of the social sciences are the same as those of natural science, though their specific concepts are distinctive and the interests the social sciences serve are sometimes different. The compromise view holds that the lack of progress in social science is a consequence of the complexity of human processes, which is much greater than that of natural processes. It also identifies limits on our understanding that stem from the regulations, mores, and inhibitions barring controlled experiments on human beings. If this view is right, the problems of social science are mainly practical instead of philosophical. Though this is a possible view, much of the work of philosophers and social scientists who have dealt with the philosophy of social science suggests that this nice compromise is a difficult one to maintain. Most of the rest of this book expounds arguments that one way or another attempt to undercut this philosophically deflationary compromise. We shall reconsider it often.

Some philosophers and social scientists reject as uninteresting or unimportant the question of whether the social sciences have progressed as fast as natural science. They hold that the question is peripheral to the philosophy of social science. On their view, the social sciences raise distinctive philosophical problems that have nothing to do with any comparison to other disciplines. On this view, the chief goal of the philosophy of social science is to understand the disciplines involved, without casting an eye to comparative questions that are at best premature and at worst a distraction.

Those social scientists and others who demand predictive improvement as the litmus test of advancement in the social and behavioral sciences condemn this attitude as complacent: it is indifferent to human needs and aspirations, which social science is called upon to serve, for the extent to which social science can ameliorate and improve the human condition is a function of its similarity to natural science as a source of predictively useful knowledge that can be applied in the way physics is applied to engineering.

There are several controversial counterarguments to this demand that social science show the sort of predictive improvement that natural science manifests and provide us with the sort of technological mastery that natural science confers.

First, this demand seems to assume that the social sciences are all of one piece, and most stand or fall together in regard to their predictive powers. It may be that some social sciences are rightly viewed as potential sources of predictive knowledge if conducted according to the "right methods." But in others, the appropriate methodology may not by any means aim at or produce this sort of technologically applicable information about human affairs. Not all the social sciences should be assessed along the same limited set of dimensions.

Second, demanding that the social sciences show persistent increases in predictive power can't make them do it. If there are any impediments to predictive success and technological application in the nature of human affairs, then no matter how hard anyone tries, predictive improvements can't happen. Third, it is often argued that the misguided belief that we already have such knowledge has been used in the past not to ameliorate the human condition but to worsen it. Even if we ever acquire such knowledge, the prospects of its beneficent use are dim. Finally, it is argued that the understanding of human affairs may ameliorate the human condition even if it does not confer on us useful tools for manipulating the social environment, however well intentioned. The sort of understanding some of the social sciences provide is precisely of this type—it enhances our lives without necessarily enabling us to control our own, or for that matter others' lives, any better.

SUMMARY

Philosophy deals with two sorts of questions, ones the sciences cannot answer and ones about why the sciences cannot answer all questions. Since the social sciences are domains of greater methodological debate, and of greater immediate relevance to daily life and to matters about which people really care a great deal, the philosophical questions about what the social sciences can and cannot tell us about human life are even more pressing than questions raised by the limits of the natural sciences.

Perhaps the leading question the philosophy of social science faces is whether we should seek the kind of understanding of human affairs that natural science gives us about nonhuman processes in nature—knowledge that enables us increasingly to predict and control phenomena.

Introduction to the Literature

The account of the nature of philosophy and its relationship to science defended here is elaborated in Alex Rosenberg, *The Philosophy of Science: A Contemporary Introduction*, in any of its three editions. This text and Peter Godfrey-Smith's *Theory and Reality* introduce the philosophy of science in two ways: Rosenberg's book is thematic, identifying the problems and alternative answers to them that philosophers of science have provided. Godfrey-Smith's book is organized historically.

The two classical introductions to the philosophy of science are Carl Hempel, *Philosophy of Natural Science*, and Ernest Nagel, *The Structure of Science*. The latter work is somewhat more difficult and far more comprehensive than Hempel's. It includes an extended defense of naturalistic philosophy of social science. The appendix to Daniel Hausman's *The Inexact and Separate Science of Economics* is a particularly clear introduction to the philosophy of science, with special relevance for the social sciences, within the compass of fifty pages.

CHAPTER TWO

The Methodological Divide: Naturalism Versus Interpretation

The main lines of dispute about how the social sciences should proceed turn on disputes in epistemology—the theory of knowledge—in particular, whether predictive success should be a necessary condition for knowledge, as in natural science, or whether we should adopt a different theory of knowledge to assess the progress of the social sciences. Which theory of knowledge we choose determines how we assess the progress the social sciences have made in understanding phenomena in their domains.

NATURALISM VERSUS INTERPRETATION

Epistemology, as noted in Chapter 1, is the study of the nature, extent, and justification of knowledge. Competing epistemologies are supposed to have implications for methodology, that is, for choosing the methods that will provide knowledge, as epistemology defines it. If the epistemology of natural science is the only correct one, then the methods of the natural sciences are the only ones that will provide knowledge in social science. If there are other epistemologies, other conceptions of knowledge, ones more appropriate for understanding human affairs, then the methods and the theories of the social sciences will inevitably differ from those of natural science.

Any comparison of progress in advancing knowledge by social and natural sciences requires an epistemological starting point: a thesis about what constitutes knowledge and how to acquire it.

First we'll outline the epistemology behind the argument that understanding is the same in the natural and the social sciences. Then we'll set out the counterargument that the comparison is based on several epistemological mistakes about both social and natural sciences. Finally we will see why all social scientists must, willy-nilly, take sides on this dispute and what it means, not just for the epistemologies of their disciplines, but for metaphysical issues raised by their decisions about epistemology.

PROGRESS AND PREDICTION

Natural science has provided increasingly reliable knowledge about the physical world since the seventeenth century. From precise predictions of the positions of the planets, the natural sciences have gone on to predict the existence and properties of chemical elements and the mechanisms of the molecular biology of life. These predictions have given weight to the increasingly precise explanations the natural sciences provide as well. In addition to systematic explanation and precise prediction, natural science has provided an accelerating application in technologies to control features of the natural world. This sustained growth of knowledge and application seems absent from the sciences of human behavior.

In social disciplines, there seem to be moments at which a breakthrough to cumulating knowledge has been achieved: Adam Smith's *Wealth of Nations*, Durkheim's work in *Suicide*, perhaps Keynes's *General Theory of Employment, Interest, and Money*, or Skinner's *Behavior of Organisms*, for instance. But subsequent developments have never confirmed such assessments. Though the social sciences have aimed at predicting and explaining human behavior and its consequences at least since the Greek historian Thucydides in the fifth century BC, some say we are really no better at it than the Greeks.

Therefore, the argument concludes, something is the matter with the social sciences; probably they are not "scientific" enough in their methods. They need to adopt methods that more successfully uncover laws or, at any rate, models and empirical generalizations, which can be improved in the direction of laws or brought together in theories that explain their applications and improve on our predictive and explanatory power when it comes to human affairs.

Why models, generalizations, and laws? It's pretty clear that technological control and predictive success come only through the discovery of general regularities, which enable us to bend the future to our desires by manipulating present conditions and, perhaps more important, enable us to prevent future misfortunes by rearranging present circumstances. The only way that is possible is through reliable knowledge of the future, knowledge of the sort that only laws can provide.

There are two other less practical and more philosophical arguments for the importance science attaches to laws. First, the kind of explanation science seeks is causal, and causal knowledge requires regularities. Second, the certification of scientific claims as knowledge comes from observation, experiments, and the collection of data. Only generalizations that bear on the future can be tested by new data.

Why does causation require laws or regularities? Consider how we distinguish a causal sequence from an accidental one. Suppose I strike a wooden match, which breaks in the middle and ignites into flame. Why do we say that the striking caused the ignition, not the breaking in half? Because match strikings are generally followed by flames, whereas match breakings rarely are. Even this regularity is not exceptional. Sometimes a struck match will not light. But to explain these failures causally and to prevent them from happening in the future, we search for further regularities, for example, that wet matches will not ignite. Our search ultimately leads to laws of chemical reactions expressed in terms that don't mention matches and their being struck. These laws have few or no exceptions and ultimately underwrite the rough generalizations that experience leads us to frame.

The eighteenth-century British philosopher David Hume was the first to argue that, independent of our past experiences, there is nothing we can directly observe in any single, observed sequence of events that enables us to detect that the first event causes the second; there is no detectable glue attaching a cause to its effects that allows us to distinguish between causal and accidental sequences. Hume's observation is reflected in the methods all the sciences have developed to distinguish causation from mere correlation: by identifying well-confirmed regularities that stand behind individual causes and that are absent in cases of mere correlation. Because strict, exceptionless laws are hard to find in everyday life, we make do with rough-and-ready empirical regularities to underwrite particular causal claims. In the sciences—natural and social—these rough regularities often take the form of statistical generalizations. Much statistical methodology is devoted to distinguishing merely coincidental statistical correlations from correlations that reflect real causal sequences.

According to Hume, when science traces observed causal sequences back to fundamental physical regularities, such as Newton's law that bodies exert gravitational attraction on one another, there is nothing more to them than universality of connection. When we reach the most fundamental laws of nature, they will themselves be nothing more than statements of constant conjunction of distinct events. In science, a causal explanation must in the end appeal to laws connecting the event to be explained with prior events. Indeed, there is no stopping place in the search for laws that are more and more fundamental. The role accorded to laws has been a continuing feature of empiricist philosophy and empirical methodology in science ever since

Hume. And the importance of generalizations, models, and other approximations to laws in all the sciences—natural as well as social—has been grounded on the role that laws play in underwriting causation.

Since, even in individual cases, our knowledge of causation is based on the preliminary identification of generalizations, which themselves are refined through the repeated observation of similar sequences, it is no surprise that such observation is what tests our explanatory and predictive hypotheses and certifies them as justified knowledge. Hume's analysis of the nature of causation as constant conjunction means that our knowledge of individual causal sequences is justified only if we can successfully predict further effects when we observe their causes. If Hume is right about causation, prediction is the sine qua non of causal knowledge.

EMPIRICISM AND LOGICAL POSITIVISM IN THE PHILOSOPHY OF SOCIAL SCIENCE

After 1900, Hume's two insights blossomed into a philosophy of science that held sway for half a century or more and set the agenda of problems in the philosophy of natural and social science. This philosophy of science was labeled by its exponents "logical positivism" and sometimes "logical empiricism" or just "positivism" for short.

Logical positivists adopted Hume's epistemology—empiricism. This is the thesis that our knowledge of the world can be justified only by the testimony of the senses—that is, by experience, observation, and experiment. Logical positivists extended this thesis to a more radical one, that theories that one could not verify or falsify by experience are, strictly speaking, meaningless. Using this principle, logical positivists stigmatized much nineteenth-century philosophy, especially the work of Hegel and his followers, who advanced theses like, "All history exhibits the self-development of reason," which seemed not only grandiloquent but also so vague that one couldn't know whether to disagree with it. Logical positivists held that such sentences were meaningless, more like, "Colorless green ideas sleep furiously," or, "'Twas brillig and the slivy toves did gyre and gimble in the wabe," than technical claims of scientific theory we can't understand till we have learned calculus and quantum theory. Positivists wanted to limit meaningful discourse to what could be tested by the methods of science and to the logical analysis of the discourse that can be provided by the methods science uses.

Thus, positivists devoted much effort to analyzing the nature of theory testing in the natural sciences. Part of their motivation for this effort was their epistemological commitment to Hume's empiricism, which accorded scientific findings the status of best-established knowledge.

Positivists also devoted much time to developing accounts of the nature of scientific theories and the structure of scientific explanation. Since scientific explanation uncovers causal mechanisms, it must involve laws. The theory of explanation they advanced is called the "deductive-nomological" or "covering-law" theory: to explain a particular event, one deduces its occurrence from a set of one or more laws of nature together with a description of the "initial" conditions that the laws require for the occurrence of the event to be explained. Thus, we can explain why a car's radiator burst by deducing from the facts that the temperature fell below freezing, and the radiator was full of water, and the law that water expands when it freezes into ice. Similarly, the positivists pointed out, given the law about the expansion of water at its freezing temperature, we can predict that the full radiator will break if we know that the temperature is falling below freezing. Thus, according to positivists, explanation and prediction are two sides of the same coin. The role of successful prediction in certifying that explanations are correct was only the beginning of its importance for the positivists' conception of knowledge.

The covering-law model of explanation can be extended to account for how science explains laws, and it can be developed into an analysis of scientific theories. Laws are explained by derivation from other, more general laws. Thus, we can derive a chemical law—for example, that hydrogen and oxygen will combine under certain conditions to produce water—by deducing it from more general physical laws governing the chemical bonds produced through the interaction of electrons. A scientific theory is just a set of very general laws that jointly enable us to derive a large number of empirical phenomena. According to the positivists, a theory has the structure of an axiomatic system—rather like Euclidean geometry with its postulates, or axioms, and its theorems derived from them by logical deduction. But unlike geometry, the axioms of a scientific theory are not taken to be known for certain. Rather, positivists held that such axioms are hypotheses, which are tested by the deduction from them of predictions about observations. If observations corroborate the predictions, the theory is confirmed to some degree. But no theory is ever conclusively verified once and for all. Theories, like laws, make universal claims. Our evidence for these claims about everywhere and always is limited to here and now and in the past. Therefore, scientific knowledge is fallible, always subject to revision, correction, improvement, as guided by its predictions that go wrong.

To emphasize the hypothetical nature of the basic laws of a theory and the logical relations between these laws and the observations that test them, positivists named their account of theories "hypothetico-deductivism." Despite the fallibility of science, the positivists (along with pretty much everyone else) held that the history of science is a history of progress, a history of increasingly powerful predictions and increasingly precise explanations of the way the world works. And positivists parlayed their analysis of the nature of theories into an account of this progress. Galileo's laws and Kepler's laws could be mathematically derived from Newton's, and Newton's from Einstein's theories of relativity—special and general, and from quantum mechanics, while all three of these could be deduced from superstring theory. The progressive accumulation of knowledge in science is thus certified by its increasing predictive success.

The history of science is the history of narrower theories being "reduced" to broader theories. One theory is reduced to another when the distinctive fundamental assumptions of the first theory—its axioms—can be derived as theorems from the fundamental assumptions of the broader theory. This will ensure that the predictive successes of a theory are preserved by its successors. Thus, Galileo's theory of terrestrial mechanics and Kepler's theory of the planetary orbits can be derived as special cases from one theory of mechanics—Newton's. And Newton's theory turns out to be a special case derivable from Einstein's special theory of relativity. Similarly, the balanced equations of chemistry follow from the physical theory of the atom, and Mendelian genetics, discovered in the nineteenth century, turns out to be derivable from molecular genetics. Or so positivists claimed. Most students of science accepted this picture of the progress of science as accumulating more and more knowledge by incorporating and preserving the predictive successes of older theories in newer ones. But the positivists attempted to make the picture precise by giving a formal account of the reduction of theories as the logical derivation of one axiomatic system from another.

As noted above, the positivists sought to parlay their empiricist theory of knowledge, which made predicting observations central, into a theory of meaning—the so-called principle of verification. This theory of meaning has its origins in the seventeenth- and eighteenth-century British empiricist notion that the meaning of a word is the image in the mind that it names. So "red" means the color experience one has looking at stop signs. The positivists developed this notion into the thesis that words have "empirical" meanings, roughly given by their roles or the contributions they make to testable sentences—whole statements whose truth or falsity can be determined by making observations, that is, predicting them and seeing whether they are borne out. Prediction is thus built into the positivist theory of knowledge, of explanation, and even of language.

One serious problem for positivists was reconciling their empiricism—the requirement that meaningful statements be testable in observation—with the unobservable entities, processes, and properties of scientific theories. It is clear that theoretically indispensable concepts like *electron*, *charge*, *acid*, and *gene* name unobserved things that we can make no observational predictions about. Are we to stigmatize the statement, "Electrons have negative charge" as meaningless because it cannot be tested? No. Positivists held that statements that use such concepts can be indirectly tested by the derivation from them of statements that are about observations. The trouble with this approach is that no particular observations follow from any single theoretical statement; experimental observations follow only from large sets of theoretical statements working together. So we cannot give the observational pedigree of a single term like "electron" or of just one statement, like "Electrons have quantized angular momentum."

The realization of this fact about the relation between observations and theory began the unraveling of positivism as a philosophy of science. If it is through their predictions that scientific claims face the court of experience, not one by one but in large groups, then we cannot distinguish the theoretical statements or individual concepts in the group that are meaningful from the ones in the group that may not be. Moreover, when observations disconfirm a set of theoretical statements that work together to imply observations that don't occur, they force us to give up one or more members of this set. But they don't point to the one we have to give up, and we can always save our favorite hypothesis from disconfirmation by making any one of a large number of other possible changes in our theory.

This "underdetermination of theory by evidence" has serious consequences for empiricist philosophy of science. Most important, it suggests that theory choice may not be governed exclusively or perhaps even largely by observation, as empiricism requires. That is because scientists' observational evidence does not confirm the theory scientists actually endorse any more strongly than it does any number of alternatives we can construct by making slight changes in the theories scientists actually believe. Accordingly, when scientists embrace specific theories, it cannot be just because those are more strongly confirmed by observation than others. It must be that the scientists' theoretical choices are driven by nonempirical factors.

What might the nonempirical factors be? By the 1970s, many philosophers of science were beginning to search for these factors. Searching for the nonempirical factors means, in effect, giving up positivism and its empiricist epistemology as an account of scientific method. In fact, it may involve giving up the positivist and empiricist claim that science provides ever increasing objective knowledge. For if the factors that govern theory choice are, say, psychological or social or ideological instead of empirical and logical, then the source of insight into the nature of science will be psychology or sociology or political science or economics or history. Certainly the only way to tell how science has actually proceeded is to explore the history of science. So, by the 1970s, the history of science had become a crucial component of any attempt to understand the nature of natural science. Soon after that, sociologists began to seek nonevidential factors that determine scientific consensus among social forces. Eventually, each of the social sciences could boast of a subdiscipline devoted to understanding the character of science and scientific change. Besides the sociology of science, there emerged, for example, the economics of science, which sought to show how scientific research reflects the rational distribution of scarce research resources in the face of uncertainty. Students of gender and gender politics sought to show that scientific practices, and in some cases scientific theories, were the result of male domination and discrimination based on race, class, and gender. These enterprises had little influence on the course of the philosophy of science itself, though they had a good deal of impact on the ways in which each of these social sciences viewed itself as a science. And all of them reduced the influence of logical positivism within the social and behavioral disciplines.

Meanwhile, for quite different reasons, logical positivism as a viable movement among philosophers had disappeared by the 1970s. This disappearance was not due to the acceptance among philosophers of views that cast doubt on science's objectivity or improvement. The eclipse of logical positivism was due much more to philosophers' own studies of the history of science, and especially sciences such as biology. What they learned was that scientific developments in these disciplines do not honor the narrow strictures of positivism. Moreover, no one could solve the problems required to vindicate its philosophical program. What former positivists and their students remained wedded to was a vision of science as objective knowledge, which, though fallible, is characterized by persistent improvement in explanatory depth, as revealed in its predictive power for observations. In its expansion and deepening of our understanding, models, general laws, and universal quantitative theories continued to be recognized as playing an indispensable role. For these philosophers and for social scientists influenced by them, the question remains why the same sort of progress in providing models, laws, and theories of ever-increasing predictive power with respect to observations—which is a feature of natural science—does not characterize the social sciences.

THE EMPIRICIST'S DIAGNOSIS: WHY SOCIAL SCIENCE FAILS TO DISCOVER LAWS

Why have the social sciences not provided increasing amounts of cumulating scientific knowledge with technological payoff for predicting and controlling social processes? The social sciences have failed, despite long attempts, because they have not uncovered laws or even empirical generalizations that could be improved in the direction of real laws about human behavior and its consequences. This diagnosis calls for both an explanation of why no laws have been discovered and a proposal about how we can go about discovering them.

One compelling explanation is that social science is just much harder than natural science: the research object is us, human beings, and we are fiercely complicated systems. It is therefore no surprise that less progress can be made in these disciplines than in ones that deal with such simple objects as quarks, chemical bonds, and chromosomes. After all, the human being is subject to all the forces natural science identifies as well as those of psychology, sociology, economics, and so forth. Teasing out the separate effects of all the forces determining our behavior is more formidable a task than that which faces any other discipline.

Add to this the restrictions of time, money, and morality on the sorts of experiments needed to uncover causal regularities, and the relatively underdeveloped character of social science should be no surprise. Perhaps the complexity of human behavior and its causes and effects are beyond our cognitive powers to understand. Perhaps there are laws of human behavior, but they are so complicated that human beings are not clever enough to uncover them. Or perhaps we haven't given social science enough time and effort; perhaps breakthroughs in, say, computer simulation will enable us to extract models, generalizations, and eventually laws from data about human behavior. On this view, the social sciences are just "young sciences." By and large, they are or can be scientific enough in their methods, but they just require more time and resources to produce the social knowledge we seek.

These explanations for the failure of social science to enable us to discover the laws governing human behavior have not convinced many students of the social sciences. Is human behavior so much more complicated than nonhuman processes? Sure, but science has always successfully coped with complexity in other cases. Are the social sciences really young by comparison with the natural sciences? From when should we date the social disciplines? From the post–World War II infusion of research money, statistical methods, cheap computation, and improved scientific education of social scientists? Should we date social science from the self-conscious attempts, like Durkheim's in the late nineteenth century, to establish a quantitative science of society? Or should we go back to the eighteenth-century Marquis de Condorcet's or the seventeenth-century Thomas Hobbes's development of rational choice theories of human behavior? Some will argue that the search for laws in social science goes back to Thucydides and the Peloponnesian War in the fifth century BC. Certainly the desire to understand and predict human behavior is at least as old as the desire to understand natural phenomena, and the search for laws of human behavior goes back past Machiavelli for sure.

For some philosophers and for even more social scientists, the claim that human behavior is too complicated to understand or that the social sciences are young rings hollow. Twentieth-century behaviorists in all the social sciences provided good illustrations of these attitudes. These social scientists offer a different explanation for the failure to discover laws. To begin with, behaviorists didn't accept the argument from the complexity of human beings to the difficulty of discovering laws about them. They note that as natural science developed, its subject matter became more complex and more difficult to work with. Indeed, to advance knowledge in physics nowadays, vast particle accelerators must be built to learn about objects on which it is extremely difficult to make even the most indirect observations. But the increasing complexity of research in the natural sciences has not resulted in any slowdown in scientific advance. Quite the contrary, if anything the rate of "progress" has increased over time. Thus, by itself, complexity can hardly be an excuse for the social sciences.

Moreover, the argument continues, the social sciences have had a great advantage over the natural sciences. It is an advantage that makes their comparative lack of progress hard to explain as merely the result of complexity and the difficulties of experimentation. In the natural sciences, the greatest obstacle to advancement has been conceptual, not factual; that is, advances have often been the result of realizing that our commonsense descriptive categories needed to be changed because they were a barrier to discovering generalizations. Thus, the Newtonian revolution was the result of realizing that commonsense notions about change, forces, motion, and the nature of space needed to be replaced if we were to uncover the real laws of motion. We had to give up our commonsense suppositions that there is a preferred direction in space, that the earth is at rest, that if something is moving there must be a force acting on it. Instead we must view motion at constant velocity as the absence of net forces, consider "down" as just the direction toward the strongest local source of gravity, and accept that Earth is moving at about seven hundred feet per second. Similarly, the pre-Darwinian conception of unchangeable, immutable species must be surrendered if we are ever to coherently entertain an evolutionary theory, still less to accept one that explains diversity by appeal to blind variation and natural selection changing old species into new ones.

But in the social sciences, the change of fundamental categories has not been thought necessary. In fact, since the very beginnings of the philosophy of social science in the late nineteenth century, it has been argued by important social scientists and philosophers that these disciplines must invoke the same framework of explanatory concepts that people use in everyday life to explain their own and other people's actions—the categories of beliefs, desires, expectations, preferences, hopes, fears, wants, that make actions *meaningful* or *intelligible* to ourselves and to one another. The reason often given for insisting on explanations of behavior that show its meaning for the agents who engage in it is that the perspective of social science is fundamentally different from that of natural science. The perspective of natural scientists is that of *spectators* of the phenomena they seek to discover. The social scientist is not just a spectator of the social domain, but a participant, an agent, a *player* in the human domain. Theories in natural science cannot change the *nature* of the reality that the physicist or chemist or biologist studies, but theories in social science can and often do. As participants in social life learn about these theories, their actions may change in light of them. This goes for social scientists as well as those whose actions and behavior they study. If laws and theories in social science must be ones that reveal the meaning of behavior and make it intelligible to the human agents who engage in it, they will have to employ the categories and concepts in which we humans have always understood our own actions and their consequences for others. Notice that those who hold this view need to give us an argument for why the social scientist cannot adopt the perspective of observer and must adopt the perspective of participant. The fact that many social scientists do so is not an argument that they inevitably must.

On the other hand, even those who seek a social science that, like natural science, provides only an observer's perspective and not that of a participant, more often than not embrace the conceptual repertoire of common sense. For what they seek are the causes and consequences of our actions, and they agree with common sense that these actions are determined by our desires and our beliefs. Accordingly, almost all social scientists have long searched for models, generalizations, and ultimately laws connecting actions, beliefs, and desires.

It is therefore hard to avoid the conclusion that the social sciences never had to face the greatest obstacle to advancement in the natural sciences: the need to carve out entirely new ways of looking at the world. Thus, we might expect progress to have been possible or perhaps even more rapid in the social than in the natural sciences. The absence of progress makes the excuse—that these are young disciplines that face subjects of great complexity—unconvincing to many social scientists and some philosophers.

In fact, behaviorists and others argue that the basic categories of social science are wrong. The reason no laws have been uncovered is that the categories of action, desire, belief, and their cognates have prevented us from discovering the laws. And many social scientists seek to supplant those categories with new ones, for example, operant conditioning, sociological functionalism, and sociobiology.

It is easy to see how a category scheme can prevent us from uncovering laws or regularities, even when they would otherwise be easy to find. Suppose we define *fish* as "aquatic animal" and then attempt to frame a generalization about how fish breathe. We do so by catching fish and examining their anatomy. Our observation leads to the hypothesis that fish breathe through gills. Casting our nets more widely, we begin to trap whales and dolphins, and then we modify our generalization to "all fish breathe through gills, save whales and dolphins." But then we start to drag along the ocean floor and discover lobsters, starfish, and crabs, not to mention jellyfish floating at the surface, all breathing in different ways. There's no point in adding more and more exceptions to our generalization. There isn't just one generalization about how all fish breathe, not as we have defined fish. The trouble is obvious: it's our definition of fish as aquatic animal. A narrower definition, such as "scaly aquatic vertebrate," will not only, as Aristotle says, "carve nature closer to the joints"—that is, reflect its real divisions more accurately—but also enable us to frame simple generalizations that stand up to testing against new data. Indeed, the difference between a "kind-term" like *gold*, which reflects real divisions in nature, and one like *fake gold*, which does not, is that there are laws about the former and not the latter. Philosophers call the kind-terms that figure in laws "natural kinds." The search for generalizations and laws in science is at the same time a search for these natural kinds.

What if the terms *desire*, *belief*, *action* do not name natural kinds? What if they just don't carve nature at the joints? Then, like our example, *fish*, every generalization that employs those terms will be so riddled with exceptions that there are no laws we can discover stated in these terms. In consequence, the explanations that employ them would inevitably have little predictive power. One solution to the social sciences' problem would be to find new explanatory variables to replace the "unnatural" kinds. Social scientists are infamous for introducing such neologisms—terms like *reinforcer* from behaviorism, *repression* from psychoanalysis, *alienation* from Marxian theory, or *anomie* from Durkheim's sociological tradition. Advocates of each of these theories promise that the application of their preferred descriptive vocabulary will enable the social sciences to begin to progress in the way the natural sciences have. If these social scientists are correct, their disciplines will indeed turn out to be young sciences. For in the absence of their preferred system of kinds and categories, the social sciences are rather like chemistry before Lavoisier: trying to describe combustion in terms of "phlogiston" instead of "oxygen," and failing because there is no such thing as phlogiston.

It is important to keep in mind that social scientists and philosophers have challenged every step in this chain of reasoning: the claim that the natural sciences show progress and the social sciences do not; the assumptions about what progress in the growth of knowledge consists in; the role of laws in providing knowledge; the purported explanations of why the social sciences have not yet uncovered any laws; and the prescriptions about how they should proceed if they hope to uncover laws. Let us examine this challenge.

REJECTING EMPIRICISM FOR INTELLIGIBILITY

Those who reject the argument that natural science has progressed and social science has languished take up their counterargument at the very foundations of the philosophy of natural science. To begin with, it is sometimes held that the natural sciences have not in fact made the kind of progress ordinarily attributed to them. In making that point, they sometimes exploit the account of science advanced in Thomas Kuhn's *Structure of Scientific Revolutions*. This book has been the most heavily cited work among social scientists writing about method and philosophy in the years since 1962 when it was published. Kuhn's book did more to undermine the dominance of logical positivism in the philosophy of natural and social science than any other single volume. It challenged several of the details of the positivist picture of science sketched in our discussion above. In particular it raised two challenges against the claim that predictive success is a universal criterion in all disciplines of progress in attaining knowledge. First, Kuhn gave reason to suppose that natural sciences do not really progress in the way orthodox history of science portrays. Second, and more radically, he challenged the role of prediction in the epistemology of natural science altogether, arguing that the demand for it was a temporary fashion of Enlightenment science. Some of Kuhn's readers, especially social scientists, interpreted Kuhn as claiming that instead of progress, the history of scientific theories from Aristotle to Einstein has been characterized by change without overall improvement. Thus the history of science is the succession of very general theories, or what

Kuhn called "paradigms," that replace one another, without making net improvements on their predecessors.

According to the opponents of the thesis that science shows cumulative "progress," the reason scientific theories do not build on their predecessors is very roughly that they constitute irreconcilably different conceptual schemes. Their notion is that scientific theories are more like poems, meaningful in one language but never adequately translated into another. Accurate comparisons between theories is impossible, for too much is lost in translation from one to another. Of course, a theory-free language to describe observations would enable us to compare two theories for predictive success if they shared the same observation language. But there is no such theory-neutral stance, and therefore one theory's confirming data will be another theory's experimental error. In retrospect, the absence of a theory-neutral language of observation can explain most of the failures of the logical positivists' program for understanding science and vindicating an empiricist epistemology. The claim that science shows persistent improvements in predictive success about what can be observed certainly becomes more controversial.

The appearance of progress in science, Kuhn held, is the result of scientists in each generation rewriting the history of their subjects so that the latest view can be cloaked in the mantle of success borne by the scientific achievements it replaces. Positivists failed to see this situation because they accepted the early twentieth-century rewrite of the history of physics as the objective truth about what actually happened in the history of science. Instead, it was just part of the new paradigm's attempt to obscure its victory over the old one. Careful study of the cases of what positivists called the reduction of narrower theories by broader ones—like the reduction of Newton's theory to Einstein's—will show that no such thing took place. Succeeding theories are incompatible with one another, so neither can be derived from the other. What can be reduced to the later, newer theory is just the newer theory's inadequate "rewrite" of the older theory.

In fact, Kuhn seemed to claim that the whole idea that predictive success should constitute a transdisciplinary criterion for scientific knowledge is part of a conceptual scheme: positivism, or empiricism, associated with Newtonian science. But this paradigm has now been replaced in physics by the theories of relativity and quantum mechanics. The conceptual schemes of those theories deprive Newtonian demands on scientific method of their authority. Newtonian science made prediction a requirement of scientific achievement because it was a deterministic theory of causal mechanisms. But quantum mechanics has revealed that the world is indeterministic; thus, definitive prediction can no longer be a necessary condition of scientific success. Nor does it make sense to search for causal mechanisms described in strict and exceptionless laws. The fundamental laws of quantum physics are statements of probabilities.

For the same reasons that scientific standards change within each of the natural sciences, they differ extensively between them and across the divide between natural and social sciences. Indeed the differences between standards in the natural and social sciences must be wider than the others. The reason for these differences is of course the vastly different "paradigms" that ground theory in each discipline. Thus, the charge that the social sciences have made less progress than the natural sciences is often said to rest on a myopic absolutism—a view that improperly generalizes from the methodological recipes of the obsolete paradigm of Newtonian physics.

Of course, within some disciplines prediction and practical application are important ways of "articulating the paradigm." But to identify what kinds of predictions, if any, are appropriate to a social science, we must first identify its "ruling paradigm." If we find the right paradigm, we will be able to see that in the light of its standards, the social sciences are progressing perfectly well, thank you. We will see that whether there is as much progress in the social as in the natural sciences is a question not worth answering. The differences between progress in physics or biology and the type of progress that characterizes the human sciences are too great even to allow comparison.

Unlike the natural sciences, which aim at causal theories that enable us to predict and control, the social sciences seek to explain behavior by rendering it meaningful or intelligible. They uncover its meaning, or significance, by interpreting what people do. The interpretation of human behavior, in this view, is not fundamentally causal. Nor is intelligibility provided by the discovery of laws or empirical generalizations of any interesting sort.

The social sciences are concerned with the part of human behavior ordinarily described as action and not with mere movements of, or at the surface of, the body. Speech, not snoring; jumping, not falling; and suicide, not mere death, are the subject matter of some of the social sciences. The parts of social science that do not deal directly with individual action—demography, econometrics, and survey research, for example—deal with actions' consequences and their aggregation into large-scale events and institutions.

Though understanding the meaning of actions is not directed at uncovering causes, it certainly satisfies some standards of predictive success: the correct interpretation of human actions enables us to navigate successfully in a society of other human beings. When we step back and consider how reliable are our predictions of the behavior of others, we cannot fail to be impressed with the implicit theory that growing up in society has provided us.

This theory, known as "common sense" or "folk psychology," tells us obvious things we all know about ourselves and others. For instance, people do things roughly because they want certain ends and believe their actions will help attain them. Folk psychology includes such obvious truths as that being burned hurts, medium-sized objects in broad daylight are detected by normal observers, and thirst causes drinking.

Folk psychology is a theory in which we repose such great confidence that nothing in ordinary life would make us give it up. That leads some people to hold that folk psychology isn't a theory, but something more fundamental, a "form of life," a way of living. After all, a theory is something we could give up; it is composed of models, empirical generalizations, or even laws that are subject to testing by experience. But when we try to express the central principles of folk psychology, we seem to produce only banal and obvious principles or ones with gaping exceptions. It's probably true by definition that people act in ways that they believe will attain their desires. And it's plainly false that thirst always causes drinking. We can dream up lots of exceptions to that generalization. If folk psychology is a theory, it's surely very different from theories in the natural sciences.

Whether it is a scientific theory or not, folk psychology is still the best theory we have for predicting the behavior of people around us, and it's the one we employ when we explain our own and others' behavior. What is more, folk psychology had already reached a high degree of predictive power well before the dawn of recorded history, long before we acquired a comparably powerful theory in natural science. Folk psychology enables us to predict by identifying the meaning of behavior—by showing that it is action undertaken in the light of beliefs and desires.

Social science, it is argued, is and should be the extension and development of this resource. It inherits the predictive strength of folk psychology. But unlike natural science, the main aim of social science is not to increase the predictive power of folk psychology. Rather, the aim is to extend folk psychology, from the understanding of everyday interactions of individuals, to the understanding of interactions among large numbers of individuals in social institutions, and among individuals whose cultures and forms of life are very different from our own.

Opponents of a "scientific" approach to the social sciences claim that much of the apparent sterility and lack of progress in these disciplines is the result of slavish attempts to force folk psychology into the mold of a scientific theory of the causes and effects of action. The social scientists and philosophers who oppose the scientific approach and those who support it agree that in certain areas the social sciences have not progressed. But the diagnosis of the former does not blame the lack of progress on the complexity of social life, the inability to undertake experiments, or the failure to find the appropriate "natural kinds." Rather, opponents of naturalism hold that many social scientists have misunderstood folk psychology, mistakenly treating it as a causal theory, to be improved by somehow sharpening its predictive power. The result, as occurred in microeconomic theory, for example, has been to produce general statements that are not laws: they are either vacuous definitions or flatly false statements. In other disciplines, such as psychology or parts of sociology, that misunderstanding has produced jargon-ridden pseudoscience.

The explanation of why parts of social science seem to find themselves at a dead end can be found here. Folk psychology has reached its maximal level of predictive power. That is because folk psychology is not a causal theory, to be improved by the kind of means scientists employ to improve theory in natural science. The predictive power of folk psychology is a sort of by-product of its real goal, which is to provide understanding through interpretation. When we accept this as the aim of social science, we will recognize the important advances it has attained. Doubts about progress will be shown to be not only groundless but also fundamentally misconceived.

Proponents of this view invite us to consider how much more we now know about other cultures, their mores, morals, institutions, social rules and conventions, values, religions, myths, art, music, and medicine, than we knew a century ago. Consider how much more we know about our own society as a result of what we have learned about other societies. Our understanding of these initially strange peoples is not the product of "scientific investigation." It is the result of the cultural anthropologist's "going native," attempting to learn about a foreign culture from the inside, coming to understand the meaning of his subject's actions in the terms his subject employs. Our understanding also reflects important discoveries about the hidden, deeper meanings behind behavior, meanings that social scientists have revealed.

This hard-won knowledge represents progress in two ways. We can understand people of differing cultures. Indeed, we can acquire as much predictive confidence about them as our own folk psychology provides us about ourselves. For what we are learning is in effect their folk psychology. Moreover, learning about other cultures teaches much about our own. Specifically, it leads us to see that what we might identify as universal or true or optimal in our beliefs, values, and institutions is really parochial, local, and merely convenient for some of us. Coming to understand another and very different society by learning the meaning of its features is a cure for moral absolutism, xenophobia, racism, and other ills. That is how social science progresses. It is not meant to provide us with the means to control the behavior of others. Indeed the understanding it provides may enable us to free ourselves from the (often unnoticed) control of others or society itself. Rather, social science is meant to provide interpretations of the actions of others and of ourselves that will enable us to place our own society in perspective.

Another thing that a scientific approach to human behavior misses, by substituting causal inquiry for understanding the meaning of human action, is the moral dimension of social science. The natural sciences aim, in part, at technological progress. That's what makes predictive power so important for them. The social sciences aim at improving the human condition. This aim entails choices the natural sciences are not called upon to make, moral choices about what will count as improvements and what will not. Making these choices requires us to identify the real meaning of social institutions, as opposed to their apparent meaning. That is why so much social science involves the critique of social arrangements and institutions as unjust to one group or another. This critique will eventually emancipate human beings from their mistaken beliefs about the meanings of social events and institutions of control and exploitation.

Influential social scientists since Max Weber have argued that theories about human behavior, human action, and human institutions need to uncover both causal laws and interpretative meanings, and that this dual requirement is what distinguishes social from natural sciences. But many social scientists have held that the conceptual apparatus we need to uncover the meaning of human events, individual or aggregate, is irreconcilable with the search for causal laws. In their view, the idea that we should replace our explanatory system with one that "carves nature at the causal joints" is based on a fundamental misunderstanding of the nature and aims of social science. One central philosophical problem for this view is very clear: What sort of conceptual confusion has led so many philosophers and social scientists down the blind alley of attempts to construct and advance a discipline that apes the wildly inappropriate methods of natural science? How could so many smart people be so wrong about what they are doing?

TAKING SIDES IN THE PHILOSOPHY OF SOCIAL SCIENCE

The two arguments summarized in this chapter cover a lot of ground, including both very practical questions of social scientific method and the most fundamental problems of philosophy. The arguments reflect polar extremes on a continuum along which most social scientists should be able to locate themselves. But though they are extreme views, these positions have real proponents. More important, all social scientists take sides on the problems the positions reflect, whether they want to or not. And that is what makes the philosophy of social science relevant to social science.

The extreme views, that social science is not scientific enough and that it is not supposed to be scientific at all, disagree on too much ever to be reconciled. No one is going to convince a proponent of either extreme that the view on the other end of the continuum is right. The reason is that the differences between them rest on very fundamental issues of philosophy, claims about epistemology, metaphysics, and ethics. Since these issues were first raised by Plato almost 2,400 years ago, philosophers have not been able to settle them.

Why should the rest of us bother about these issues? They cannot be settled, and we don't occupy the extreme positions in the philosophy of social science. For these reasons, many social scientists aren't interested in the subject at all. They seem to have a good reason not to be, if the problems of philosophy are insoluble. And yet, however insoluble these problems may be, they certainly are not irrelevant.

Between the polar extremes in the philosophy of social science, there are many intermediate theories of the nature of social science that seek to reconcile the social sciences' differences from natural science with the demand that they be truly scientific in the natural scientist's sense of the term. But partisans of the extreme views agree with one another that such compromises are in one way or another incoherent—attempts to have one's cake and eat it, too. In trying to give the differences and the similarities between the sciences—social and natural—their due, the compromises turn out to be contradictory or inconsistent, or just plain false. In philosophical matters, the policy of finding a happy medium that splits the difference between rival theories is often impossible, for the positions are logically incompatible. Picking and choosing components of these two philosophies, with a view to developing a "third way," may result in an incoherent position. Or if the position is coherent, the resulting theory may not be strong enough to withstand the arguments advanced by proponents of one or both of the extreme positions.

For example, economists and political scientists are committed to explaining action in terms of the quantitative models of "rational choice theory," according to which individual expectations and preferences cause human actions. These social scientists need to explain why we have secured no predictively reliable laws about the causes of individual action. It won't do simply to say that rational choice models are idealizations or approximations like those of, say, physics—approximations that will eventually be improved in the direction of laws. For ideal models in physics have what economic models conspicuously lack: great predictive power. In addition, physical, chemical, and biological models seem capable of refinement in ways no rational choice theory has yet manifested.

Of course, economists might try to show why no regularities governing human action are necessary to explain economic choice causally. Such an argument would in effect claim for economic processes a sort of causality unknown in natural science. That causality would need a significant and unavoidably philosophical explanation. Alternatively, economists could adopt the view that the knowledge they provide is not causal but, at best, information that helps us interpret the actions of consumers in late capitalist society. Anyone acquainted with modern economic theory will recognize that view as unacceptable to economists. But unless they can provide some philosophical underpinning for their theory, economists are vulnerable to the charge that their explanatory variables are not natural kinds. That is, the explanatory variables of rational choice theory need to be surrendered in any serious causal theory of human behavior. In effect, finding an intermediate position for economics involves facing several classical metaphysical problems about causation.

The sociologist or cultural anthropologist faces a different sort of philosophical problem. Anyone who brings back an account of the meanings of actions, rituals, or symbols of other cultures must assure us that the account is correct, that it constitutes an addition to knowledge about the culture reported. How can we tell whether it is information or misinformation? This question becomes especially pressing if we, like some anthropologists, reject the demand that our theories about cultural meanings identify causes and have predictive consequences we can test. Social scientists who reject improving predictive success as a mark of knowledge have taken sides willy-nilly in the most profound disputes of epistemology. For certification as knowledge by means of observed predictions is the touchstone of empiricism. The only alternative these social scientists can adopt is some version of rationalism, the epistemology according to which at least some of our knowledge is justified independently of experience, a priori.

Social scientists who wish to embrace the natural scientific approach to human behavior often also hope to learn from their research how morally to better the human condition. They must face several of the thorniest problems of moral philosophy. First, they must show how to derive what ought to be the case—the moral improvement—from what is the case, as revealed by social research. This is a derivation widely held to be impossible by philosophers. What is needed is nothing less than an explanation of how we can acquire moral knowledge "scientifically." Moreover, if acquiring moral knowledge is possible, they must show why such knowledge does not justify paternalistic imposition of its particular claims on a potentially unwilling society.

NATURALISM VERSUS INTERPRETATION

Most empirical social scientists believe that prediction and interpretation can be reconciled. They believe that there is a causal theory of human behavior and that we can uncover models, regularities, and perhaps eventually laws that will enable us to predict human action. Let us call such social scientists "naturalists" to indicate their commitment to methods adapted from the natural sciences. Naturalists believe they can endorse the methods of natural science while doing justice to the meaningfulness and significance of human action. Thus, they do not think anything can force them to choose between these two commitments.

Naturalism's pursuit of reconciliation of prediction and interpretation has been subject to repeated objection over the course of the past hundred years. Many current controversies about social science are but reiterations of this objection and replies to it. Prominent social scientists, historians, philosophers, and cultural critics have held that we cannot do justice to actions as meaningful while at the same time seeking a naturalistic or scientific explanation of them. These critics of naturalism hold that the aim of the social sciences must be interpretation, and this means they cannot be experimental, empirical, or predictive sciences. They have adopted a succession of labels since the late nineteenth century: idealists, phenomenologists, structuralists, ethnomethodologists, and students of semiotics, hermeneutics, postmodernism, and deconstruction. These views share a rejection of naturalism and a commitment to interpretation. Therefore, we may refer to their views as antinaturalism or interpretative social science.

The history of science presents both naturalists and antinaturalists with the same problem. Prehistoric civilizations explained all natural events—especially catastrophes—in terms of the purposes of supernatural agents. Today, religions continue to do so. In each of the revolutions in Western science, the greatest obstacle to scientific advance has been the conviction that only purposes or meanings that made things intelligible could really explain them. The history of natural science is one of ever-increasing explanatory and predictive power. Science has achieved that by successively eliminating meaning, purpose, or significance from nature. After Galileo, the stars and planets were deprived of the goals Aristotelian science attributed to them; then Newton showed that force, acceleration, and gravitational attraction were enough to explain all motion. Eventually Darwin showed that the fitness of flora and fauna to their environments was to be explained without attributing purpose to them or intentions to their creator. Contemporary molecular biology has revealed the purely chemical underlying mechanism for all the biological processes that seemed originally to be explained by the goals they seek. Now the only arena in which explanations appeal to purposes, goals, intentions, and meaning is their "home base," human action.

The record of the history of science requires every social scientist to face the question, Why should human behavior be an exception to this alleged pattern? Why should meaning, purpose, goal, and intention, which have no role elsewhere in science, have the central place they occupy in social science? The obvious answer is that people, unlike (most) animals, vegetables, and minerals, have minds, beliefs, desires, intentions, goals, and purposes. These things give their lives and actions meaning, significance, make them intelligible. But what is so different about minds from everything else under heaven and earth that makes the approach to understanding people so different or so much more difficult than everything else? Every potential answer to this question is general enough, metatheoretical enough, and abstract enough to count as an exercise in the philosophy of social science.

Introduction to the Literature

Among introductions to the philosophy of social science, the best of the earlier generation of texts is Alan Ryan, *Philosophy of Social Sciences*. Ryan has also edited an anthology on the subject, *The Philosophy of Social Explanation*. Michael Martin and Lee McIntyre, eds., *Readings in the Philosophy of Social Science*, is by far the most complete anthology of influential papers written in the past generation on the subject. It supersedes two older anthologies that include many papers discussing topics treated in this book, L. I. Krimerman, ed., *The Nature and Scope of Social Science*, and M. Brodbeck, ed., *Readings in the Philosophy of Social Sciences*. Introductions to the literature in subsequent chapters will identify relevant papers in Martin and McIntyre. Papers in Martin and McIntyre of special relevance to this chapter are F. Machlup, "Are the Social Sciences Really Inferior?" and M. Scriven, "A Possible Distinction Between Traditional Scientific Disciplines and the Study of Human Behavior." Steel and Guala's recent *The Philosophy of Social Science Reader* reprints many important articles. Introductions to the literature in future chapters will identify appropriate readings from Krimerman, MacIntyre and Martin, and Steel and Guala.

The locus classicus of the naturalistic view of method in social science is David Hume, *Treatise of Human Nature* and *Enquiry Concerning Human*

Understanding. In these works one will also find the account of causation as law-governed that motivates much empiricist philosophy of science. Hume's approach is in many ways developed in John Stuart Mill, *A System of Logic*, especially Book 6, "On the Logic of the Moral Sciences." Among sociologists, the most famous defense of a naturalistic position is Emile Durkheim's *The Rules of the Sociological Method.* David Papineau, *For Science in the Social Sciences*, and David Henderson, *Interpretation and Explanation in the Human Sciences*, advance recent arguments for naturalism. Robert Brown, *The Nature of Social Laws*, provides a concise account of the history, from Machiavelli to Mill, of the claim that there are such regularities. In *Science and Human Behavior* B. F. Skinner carries the argument for this approach to social science further than any other social scientist has. Among philosophers, a different argument with similar conclusions is provided in Paul Churchland, *Scientific Realism and the Plasticity of Mind.*

The antinaturalistic view has closely embraced T. Kuhn's *Structure of Scientific Revolutions* as a counterweight to postpositivist philosophy of science. An anthology tracing the influence of Kuhn's book in social science is G. Gutting, ed., *Paradigms and Revolutions.* Interpretive social science is a tradition that goes back to the nineteenth century. Its history is traced and defended in R. G. Collingwood, *The Idea of History.* Among social scientists, its locus classicus is by Alfred Schutz. Papers by Schutz defending this view are anthologized in M. Natanson, ed., *Philosophy of Social Science*, and David Braybrooke, ed., *Philosophical Problems of the Social Sciences.* An influential work elaborating on this view is P. Winch, *The Idea of a Social Science.* A. Giddens, S*ociology: A Brief but Critical Introduction*, sketches a position that brings this approach together with critical theory (see his Chapter 4, "Critical Theory"). More recently, a radical version of interpretativism has been championed by Richard Rorty, *Contingency, Irony, and Solidarity.* For a cogent introduction to recent arguments for interpretativism, see James Bohman, *New Philosophy of Social Science.* D. Braybrooke, *Philosophy of Social Science*, attempts to reconcile the two traditions, naturalism and antinaturalism, as does Daniel Little, *Varieties of Social Explanation.* Both of these books are recommended as somewhat more intermediate texts in the philosophy of social science. M. Hollis, *The Philosophy of Social Science*, provides a powerful but difficult treatment of the competition between four sorts of research programs in social science: the structural versus the individual, and the causal versus the interpretive. It is strongly recommended for students with further interests in the philosophy of social sciences.

CHAPTER THREE

The Explanation of Human Action

The social sciences seek to understand human actions and their consequences. This makes it difficult to do justice to the demand for improvements in predictive power. The reasons have nothing to do with free will. They have much to do with the kind of explanations we want in social sciences, explanations that provide interpretations of human actions in terms of the desires and beliefs that lead to them.

ACTION, NOT BEHAVIOR

We can divide human activities roughly into two classes: "mere" behavior and action. Mere behavior includes what happens inside our bodies, such as the beating of our hearts, or at the body's surface, reflexive withdrawal from painful stimuli, or the opening and closing of the eye's iris. Action differs from mere behavior. It is what we do as opposed to what happens to our bodies. Actions are behaviors that are somehow under our control or could be, if we gave them enough thought. The difference between action and behavior is difficult to state. Some things we do seem to fall on the borderline between these two categories—yawns, for example. And sometimes actions and mere behavior are hard to tell apart: there could be no difference in the body's movement between a blink and a wink. But the difference between them is crucial for the social sciences.

Social science begins with the aim of explaining human action—not mere behavior. When and why the eye blinks is a matter for physiology, not social science. But when and why people wink at others is a question that does concern the anthropologist, the sociologist, and the psychologist. This is because winking is an *action*. Social science begins with the objective of explaining action, but it does not end there. Much social science is concerned with explaining large-scale events, for example, inflation or war. It is also concerned with institutions, such as the jury system or marriage rules. Social scientists also try to uncover and explain statistical findings about large groups. But the large-scale events, social institutions, and statistical regularities are made up of organized aggregations of individual actions.

Some social scientists, especially psychologists, have been unhappy with a hard-and-fast distinction between action and behavior. They have sought to explain human action by showing that it is nothing but "mere" behavior, albeit more complex than blinks and twitches. Other social scientists have no interest in explaining what single or small numbers of people do—regardless of whether it is action or behavior; still other sociologists and economists hold that their disciplines should focus only on large-scale social phenomena, not what single or small numbers of people do. In Chapters 8, 9, and 10 we examine these arguments. But because the action/behavior distinction is so deeply entrenched in our conception of ourselves as human, it takes a profound and powerful argument to undermine social science's interest in explaining what people do as human action.

To see why some social scientists are tempted to surrender an interest in explaining action, we need to see how action is explained in ordinary life and in social science. Long before the self-conscious attempts of the social scientist, common sense had provided us all with a theory about the behavior of our fellow human beings. It is a theory we use every day to form our expectations about the behavior of others and to explain to others our own behavior. This implicit theory is what we labeled "folk psychology" in Chapter 2. It has always been the natural starting place for explanations social scientists have given.

In fact, some social scientists, like historians, explain human actions without any explicit theory at all. Although often unaware that they are doing so, they employ folk psychology as an implicit theory for the events they explain. Of course, many social scientists have been well aware of the commonsense theory they were endorsing and have made great efforts to improve it. Microeconomic theory is perhaps the best example of this approach, and we shall consider its improvements on common sense in Chapter 6. But first we need to identify the commonsense theory and consider how it works, what kind of a theory it is, and what sorts of explanations it provides.

We will discover that there are serious obstacles facing any effort to treat folk psychology as a theory of the sort we are familiar with in natural science, or even as a first approximation to such a theory. The consequences of this difficulty for the view we have called naturalism are serious and troubling.

MAKING FOLK PSYCHOLOGY EXPLICIT

We explain human actions by identifying the beliefs and desires that lead to them. Often, such explanations are elliptical or abbreviated or proceed by means of tacit assumptions that most people can be expected to share. Suppose we explain why someone moved her king in a game of chess simply by saying she did it to avoid being in check. The explanation works because it assumes that whoever seeks the explanation knows a fair amount about chess and about human motivation. The explanation also assumes that the player wanted to avoid check and believed that moving her king was a way of doing so. Usually we don't mention the desire to avoid check because it is obvious. We don't even say there is a desire. We simply attribute a purpose (to avoid check). And we don't bother to make it explicit that the player believed that moving the king would attain this purpose. Those features of commonsense explanations "go without saying." We don't need to make all the assumptions explicit, because they are part of a form of explanation even children understand by the time they are four or five years old. In fact, the form is so well understood that it is often difficult to self-consciously expose all its assumptions, presuppositions, and implications.

We can find the same form of explanation in many a classic of social scientific explanation. For example, in *The Protestant Ethic and the Spirit of Capitalism*, Max Weber sought to explain why capitalism emerged first in western Europe, and not elsewhere in the world—for example, in India and China, where levels of population, wealth, and trade appeared to be more favorable to commerce. Weber found the critical factors present among Europeans and absent among Asians to be a peculiar combination of beliefs and desires, ones that do not directly motivate or identify the means to the accumulation of wealth.

First Weber noted that attitudes fostering capitalism appeared earliest in the seventeenth century and eventually flourished among certain Protestant sects, particularly Calvinist denominations, ones excluded from the political, military, and ecclesiastical power of the Catholic Church and nobility. Then he sought to identify the specific beliefs and desires of these seventeenth-century Calvinists that turned them toward a new mode of economic organization, one characterized by persistent investment of wealth and not consumption. Calvinism is distinctive in its commitment to predestination: God's omniscience means that he knows which individuals are to be saved and which damned long before their birth, indeed at the beginning of time. Accordingly, there is nothing individuals can do to ensure their fate after death. Of course Calvinists had an intense desire to know their predestined fate as saved or damned. This led to a belief that worldly success may be a sign of salvation, especially when the successful person was indifferent to the material rewards and did not consume them, but instead sought to employ them efficiently in ways that would reassure the person that he or she had indeed already been elected for salvation. The combination of the desire to know one's predestined fate and this belief about how to attain that knowledge make intelligible to us their commitment to "rational accumulation" in the face of traditional objections to it. Weber's explanation for the greatest change in western European history from the Renaissance to the present exploits a folk psychological theory he does not even need to articulate explicitly.

Even psychologists have adopted folk psychology as the first step in scientific psychological theory. Freud sought to substitute a deeper psychoanalytical explanation of our actions for the one common sense provides. Nevertheless, he adopted its commitment to desires and beliefs as the determinants of behavior. However, in Freud's theory, the determinants are unconscious, repressed desires and neurotic beliefs. They explain our actions, which, according to Freud, we don't even really understand ourselves.

Therefore, we explain an action by identifying the desires and beliefs that give rise to it. It is typical of philosophers to go beyond this point—on which all agree—to ask why. That is, why does citing the agent's desire and belief explain the agent's action? What connection obtains between the desire and belief and the action they together explain that makes them relevant to the action? Why does the connection between them satisfy our curiosity about why the action happened? It is hard to have patience with this question. It seems so obvious that citing the desires and beliefs explains the actions that nothing further need be said or could be said, for that matter. In fact, there is a great deal to be said about why and how desires and beliefs explain actions.

Naturalistic philosophers and social scientists have one set of answers to this question. Interpretationalist philosophers, historians, and interpretative social scientists have another. Naturalists hold that if we can identify the link between beliefs and desires on the one hand and action on the other, then we will be able to improve upon folk psychology's explanations of human action. Their opponents will argue that only by identifying this link can we convince misguided social scientists that understanding human action is a matter of meaning and intelligibility, not a matter of causation. Both sides to our dispute about the progress of the social sciences agree that we cannot ignore the question of what enables beliefs and desires to explain actions.

Will the connection among desires, beliefs, and actions be more like one described in a mathematical truth or more like a general law? Mathematical truths, such as $4^2 = 16$, state connections that exist in virtue of definitions; we can "see" their truths or work them out for ourselves, once we know what the concepts of *number*, *exponent*, and *equals* mean. These truths express connections among concepts that are intelligible to us. General laws express connections, too: for instance, copper melts at 1,083°C, but this connection is not one we could work out just knowing the meanings of the concepts *copper* and *melting*, and the centigrade scale. We need to conduct experiments that reveal the causal connection between copper's change of state and measurements of its temperature.

We should ask whether the explanatory connection between desires, beliefs, and actions will be different from either mathematical or empirical ones. On our answer will hinge the proper research strategy for social science: Should it be more like natural science, focused on observation and experiment to discover laws; or more like mathematics, focused on logical connections and the meanings of its concepts; or should it be different from both?

Go back to the explanation of why the chess player moved her king. Consider why you took your umbrella today after checking the weather forecast. Why did Hitler invade Russia? Why do young self-employed males decline to purchase health insurance? Why do white southerners in the United States tend to vote for Republicans? The answers to these questions offered in ordinary life, in the analysis of chess matches, and in history, economics, and political science, all share an explanatory theory. Let's try to extract it and state it explicitly.

Something like the following oversimplified general statement seems to lie behind our ordinary explanations of human action, and the explanations in social science that trade on folk psychology:

> [L] If any person, agent, individual, wants some outcome, *d*, and believes that an action, *a*, is a means to attain *d* under the circumstances, then *x* does *a*.

[L] is an if . . . then statement in which we fill the if-part—the antecedent—with what the person wants and believes, and in which we fill the then-part—the consequent—with a description of what the person does. Other equivalent ways to express [L] are, "Whenever an agent wants *d* and believes *a* is a means to attain *d*, then the agent does *a*," or "All agents who want *d* and believe *a* is a means to attain *d*, do *a*."

[L] is supposed to be a general statement that is true about every human being's actions, and not just ones around now. It makes a claim about the actions of the earliest Australian Aborigines, Ancient Greeks, ninth-century Norse Vikings, eleventh-century Arab traders, contemporary Kalahari Bushmen, Samurai warriors in the seventeenth century, even other hominid species, like Neanderthals, so long as they engaged in complex voluntary actions.

[L] looks like a law of the sort familiar from the sciences, such as, "If an object is in free fall, it has constant acceleration," or "Whenever metals are attached to electrical sources, they conduct current," or "All copper melts at 1,083 degrees centigrade." [L] looks like a law, but it may not be a law, as we shall see.

We may not realize that our ordinary explanations in everyday life are underwritten by [L]. Nor, for that matter, is it often recognized that much history and all biography assumes everyone accepts something like [L] for their explanations of what people do to actually provide understanding. Even much anthropology, sociology, and political science helps itself to the use of [L] without even mentioning it. As we will see, the only social science that is always explicit about [L] and its explanatory role is economics. One reason almost no one is explicit about [L]'s role in explanation in ordinary life, history, or social sciences is that when you think about it, [L] is obvious, and learned so early in life that it's hard to notice its role in explaining our own or other people's actions. In fact, it seems a bit pretentious to call [L] a "law" and describe it as part of a theory.

Nevertheless, if [L] stands behind our explanations of human action, making explicit why they came about, then we are treating [L] as something like a law of nature, part of a general theory about human behavior and action. That's why this book uses the letter L, from *law*, as the label for [L].

[L] may be the leading principle of folk psychology, but it isn't the only one. There are others, ones that connect our desires and beliefs to our environments and to our past experiences. For example, there's the "generalization" that under normal conditions we can see medium-sized objects placed before our eyes. Or again, if someone has been deprived of water for several hours, then other things being equal, the water-deprived human desires to drink. There are, of course, many more such generalizations about our perceptual abilities, beliefs, desires, memories, fears, hopes, regrets, and so forth embedded in folk psychology. But as we shall see, the most central of them to our explanatory needs in social science is [L]. For it is [L] that connects what happens around us, in the circumstances of our environment, to the movement of our bodies in actions (including speech). [L] makes this connection by identifying our desires about how we want to rearrange our circumstances, and our beliefs about how our circumstances can be shaped to satisfy our desires, to narrow down the range of actions that will attain the desires in light of the beliefs.

Like other generalizations of folk psychology, [L] must be understood either as embodying a pretty strong "other things being equal" (or ceteris paribus clause). That is because, as it stands, [L] is false. To see the falsity, consider how easy it is to construct exceptions to [L]. Suppose that *x* wants *d*, but *x* wants something else, *d'*, that conflicts with *d*, even more strongly than he wants *d*. Then *x* won't do action *a* even though *x* wants *d* and believes that *a* is the way to attain *d*. If two quite different actions, *a* and *a'*, are equally good means to attain *d*, *x* may not undertake action *a* but instead undertake action *a'* to attain *d'*. Or suppose *x* believes that *a* is a means of attaining *d*, but not the best, most efficient or enjoyable, or cheapest means of attaining it. Then *x* won't do *a* even if he wants *d*, has no overriding incompatible wants, and believes *a* is a means of attaining *d*. Things can get even worse for [L]. Even if *x* believes that *a* is the best means of securing *d*, *x* may not know how to do *a* or, knowing how, may be unable to do it. There are lots of other ways [L] can go wrong.

We can, of course, "improve" [L] to avoid all these problems, by adding clauses to it covering each of them. The result will be a much more complex statement like this:

For any agent *x*, if

1. *x* wants *d*,
2. *x* believes that doing *a* is a means to bring about *d* under the circumstances,
3. there is no action believed by *x* to be a way of bringing about *d* that under the circumstances is more preferred by *x*,
4. *x* has no wants that override *d*,
5. *x* knows how to do *a*,
6. *x* is able to do *a*,

then

7. *x* does *a*.

This is quite a mouthful, but even the four clauses we have added to [L] still may not be enough. An acute philosopher could doubtless construct a counterexample, a case that shows that the first six conditions are not sufficient for someone's doing *a*, so that we need to add a seventh, and perhaps an eighth, et cetera.

Instead of adding clauses to [L], we could simply treat [L] as bearing an "other things being equal" clause that implicitly excludes each of the exceptions covered by (3) through (6) and any further clauses [L] needs. But when, one wants to ask, can we be sure other things are equal, so that [L] can explain why the desire for *d* and the belief about *a* lead to the doing of *a*? When can we be sure that other things are not equal, so that the desire's and the belief's failure to produce the action doesn't undermine our confidence in [L] as a causal law? An unremovable ceteris paribus clause, requiring that other things be equal for the desire for *d* and the belief about *a* together to lead to *x*'s doing *a*, opens up [L] to a potentially serious charge of vagueness. This is a point to which we shall return.

It's worth noting right away that economists avail themselves of [L] under the label "rational choice theory." When they do so, their version often includes all the provisions mentioned in the expanded version. Microeconomic explanations of, say, consumer choice attribute to economic agents perfect information about available alternative choices, complete information about the constraints within which the consumer operates, and a consistent preference order—a ranking of wants—that lead the consumer to a unique choice. Though the theory is expressed in terms like *expectations* and *preferences*, they are just cognates for the beliefs and desires that figure in [L]. We'll return to economics and the role [L] plays in it in Chapter 6.

REASONS AND CAUSES

We have identified a general statement (with or without a ceteris paribus clause) that connects beliefs and desires to actions and thus can serve to underwrite our explanations, both ordinary and social scientific. What kind of a general claim is it? [L] certainly looks like a statement that identifies the causes of actions, and thus bids fair to be a law about human action, or at least to be an important approximation of a law. And that is just what the naturalist needs to vindicate a scientific approach to explaining action. Scientific explanation is causal. Therefore, any scientific approach to explanation in social science should attempt to establish a causal connection to underwrite its explanations. That will be the function of [L]. It is a causal law or a good approximation of one.

This naturalistic approach to the relation between folk psychology and a science of human action has long been associated with the views of Max Weber. Weber insisted that the sort of explanation that desires and beliefs provide must be like those provided in the explanations characteristic of natural science. Weber viewed a general statement like [L] as an "ideal type"—an unrealistic model. Like models in natural science, [L] needs to be filled in and refined in its application to individual actions. (The role of ideal types and models generally is discussed in the "Instrumentalism and Modeling in Economics" in Chapter 6 and throughout Chapters 11 and 12.) But what is crucial to Weber was the insight that scientific explanation requires laws, or approximations to them that we can eventually improve into laws.

However, as Weber recognized, unlike (other) causes, beliefs and desires are also *reasons* for actions: they justify them, show them to be rational, appropriate, efficient, reasonable, correct. They render them intelligible. So, perhaps explanations in ordinary life and social science work by showing that actions are reasonable, efficient, appropriate, or rational in the light of the agent's beliefs and desires. In this case, [L] will certainly not work like a model, a regularity, or the precursor to a causal law. For causal laws don't provide "intelligibility." Rather, [L] will reflect the fact that beliefs and desires justify or underwrite some action as reasonable. If that is how [L] works, then the fundamental explanatory strategy in social science is not that of revealing causes and effects. The aim is, instead, to make the action rationally intelligible. If [L] also mentions causes, perhaps that is a by-product of its role in making actions intelligible.

In fact, [L] has often been identified as a defining mark of rationality: an agent is rational to the extent that he undertakes the actions that are best justified, given his ends—that is, his desires or wants. Thus, far from being a contingent law describing the causes of actions, [L] turns out to be true by the definition, implicit or explicit, of what it means to be a rational agent. On this view, the social sciences that exploit [L] are not inquiries into the causation of various actions. Rather, the social sciences are investigations into the degree to which people's behavior reflects the actions of a rational agent as defined by [L].

The difference between reasons and causes is crucial, and every account of the explanation of human action must face it. The difference between them is sometimes difficult to keep clear, especially if, as most social scientists hold, beliefs and desires are at the same time both the reasons for actions and their causes.

But if they are both, then why distinguish between reasons and causes? We make the distinction because we need to identify where the explanatory power of action explanations lies. Even if beliefs and desires are both reasons and causes, their explanatory power with respect to actions may rest on only one of these two features. And on which of them it rests will determine much about the methodology of social science. If reasons for actions explain because they *bring about* actions, then the naturalism described in Chapter 2 is vindicated: the social sciences must search for causal models, generalizations, and laws; if the causes of actions explain because they are reasons for actions, then the aim of science is interpretation and intelligibility, and the antinaturalist approach to social science turns out to be correct.

So, what's the difference between reasons and causes? It's more easily illustrated than expressed. Let's consider three cases in which both reasons and causes are present. In the first, the causes explain the action and the reason does not; in the second, the reason is the cause, but the explanation is not causal; in the third case, there is a hidden reason that does the explaining. The three cases should make clear how causes and reasons differ in their explanatory roles.

Suppose we ask a jogger why she runs ten kilometers a day. She replies, "Because it's good for me." There is a fair amount left unsaid in this typical explanation. First, it's not just that running is good for her, it's that she believes it is. Second, she wants to do things that are good for her. Third, she believes that jogging won't prevent her from doing other things equally good for her. Doubtless there are other things she wants and believes that are "understood" in this explanation. All these things justify her jogging ten kilometers a day. They make it seem intelligible, reasonable, rational: if we were in her shoes, that is, had her beliefs and desires, we'd jog that much, too.

But suppose the "real reason" she jogs every day is physiological. Suppose that unnoticed changes in her body have over the years addicted her to it, so that if she doesn't jog, she feels lousy all day. Though she never notices this correlation, it keeps her running by physiologically punishing her for skipping a day and rewarding her with a "runner's high" when she does run. This is a typical explanation of a behavioral psychologist. In this explanation, there are quotes around "real reason" because the physiological facts aren't reasons at all. If they explain her jogging, then clearly they do not do so by justifying her running, making it reasonable, but by causing it in virtue of some contingent causal law she is not even aware of. In this case, reasons do not explain behavior; causes do.

But now compare the case of the runner free of this physiological compulsion, one who really does run because she believes it is good for her, believes it most efficiently fulfills her desire to stay healthy in the light of her other beliefs about how to do so. Presumably, her beliefs about running and her desire to remain healthy are stored in her brain. As such, these beliefs and desires are physical states that cause her running in accordance with as yet unknown laws of neurophysiology. We do not know these laws, but we don't need to. They are irrelevant to our explanation of why she runs. Our explanation works because the desire to be healthy and the belief that running contributes to health *justify* running. They make it *rational* in their light. They enable us to see that we would want to run if we had the same desires and beliefs. In short, these reasons explain because they make the action *intelligible* to us. We could reconstruct the reasoning that leads from the belief and the desire to the jogging in an argument, a chain of reasoning:

Belief: If you want to stay healthy, then jog.
Desire: You want to stay healthy.
Therefore
Action: Jog.

These logical connections are not to be found in causal sequences. It is these connections that distinguish reasons from mere causes.

Consider a third case, one more interesting to the social scientist. Imagine a woman who hates the smell of cigarette smoke and claims that secondary cigarette smoke is harmful. Suppose this person believes that there is no scientific evidence for her claim. If asked to explain why she thinks cigarettes are harmful to nonsmokers, the woman claims to have beliefs about evidence that secondary smoke is harmful. But what if her "real reason" for saying that cigarettes are harmful is her hatred of cigarette smoke's smell, which she has no reason to think harmful. How does this "real reason" explain her claim? It does so via [L] or something like it: a principle to the effect that it is reasonable to do the things that one believes will lead to the attainment of one's desires. In this case it leads one to say that cigarette smoking is harmful because the statement may reduce the smoking of others and thus reduce the woman's exposure to the smell she hates.

Much social science is devoted to ferreting out the "real reasons" from the ones people offer to explain or excuse their actions. But for us the question is, How do real reasons explain the actions they bring about? Is it because of the logical argument we can use to reconstruct the real reasons that justify people's actions, or is it because real reasons cause their actions? Does the action-justifying character of reasons make them a special subclass of causes, which in the end work like other causes, or does it make reasons a different sort of explainer entirely?

That is a question about [L]. For [L] is what connects people's real reasons, their beliefs and desires, to their actions. Does [L] underwrite our explanations of actions because it describes causal relations—that is, lawlike connections—in virtue of which actions are brought about by beliefs and desires? Or does [L] underwrite these explanations because it helps us identify the reasons that make a particular action justified, intelligible, rational, meaningful, or somehow significant to us?

It would perhaps be neatest and simplest if [L] does both: helps us identify the causes for actions and the reasons for them. It would thus be the key to both the prediction of much human behavior and the intelligibility of all human action. That is the view that Weber and his successors among the naturalists have embraced. But as we shall now see, this happy reconciliation is difficult to defend.

CAN REASONS WORK LIKE CAUSES?

Chapter 2 identified a distinctive problem of the philosophy of social science to be that of explaining or explaining away the alleged lack of progress of the human sciences by comparison to natural science. This problem has haunted the social sciences largely because it has vexed folk psychology's theory of human action. The problem of progress is actually two: (1) the problem facing attempts to improve [L] into a general theory of human behavior with increasing explanatory unity and predictive precision; and (2) the problem of why, in the absence of improvement of [L], no replacement for it has been found—or indeed sought. To see what these problems are, let us try to apply [L] to the causal explanation of a particular action. In doing so we need to recall what Hume's analysis of causation showed us about the connection between causation, generalization, and prediction. Our knowledge of what causes particular individual events requires us at least implicitly to frame hypotheses that have some generality, and to confirm those hypotheses by successful prediction. This process need not reach the level of physical law to provide causal knowledge. But to count as knowledge of causes, our beliefs about causes must have some predictive consequences borne out by experience.

There is a standard "recipe" given in the philosophy of science for causally explaining an individual event, the "deductive-nomological," or "covering-law," account: the occurrence of the event should be derivable from one or more general laws and a statement of "initial" conditions. The latter are roughly the set of circumstances or conditions that constitute the cause of the event to be explained. For example, we can explain why the gas in a certain container maintains a given pressure by deriving that pressure from the ideal gas law, PV = nRT, and a statement about the temperature of the gas and the volume of the container. This model is called deductive-nomological because it is a deductive argument from the initial conditions and a law that "covers" the events to be explained—connects the initial conditions to the explained event as their effect. Note that the information in a deductive-nomological explanation of the gas's pressure would have just as well enabled us to predict it as to explain it.

Of course, natural science isn't generally interested in explaining particular events, and that is true to a lesser extent of many of the social sciences. Only history is explicitly devoted to explaining particular events. But such explanation is important because it provides a means of testing and improving the laws and theories that must be employed in explaining these events.

Scientifically explaining a particular human action presumably involves deriving a statement describing the action from [L] and a set of statements about the agent's desires and beliefs. Let us take a prosaic example: Smith is carrying an umbrella as he goes to work. Why? Here is an explanation:

Initial conditions:

1. Smith wants *d*, to stay dry today.
2. Smith believes that *a*, carrying an umbrella, is the best way for him to *d*, stay dry today.

Law:

3. For any agent *x*, if *x* wants *d* and *x* believes that doing *a* is the best way for him to secure *d*, then *x* does *a*.

Therefore:

4. Smith does *a*, carries an umbrella today.

This explanation may be stilted. But it is what stands behind the briefer explanation, Smith thinks it's gonna rain. Now, how good a scientific explanation is it?

Suppose it were challenged. What if we were to demand evidence to show that the initial conditions actually obtained. For all we know, Smith might always carry an umbrella, rain or shine. He would do so if he were a British merchant banker, who always carries an umbrella because it's part of the required uniform. Or maybe Smith thinks he needs a cane to walk but doesn't want to look like an invalid, so he uses an umbrella. It's even possible that Smith wants to get wet today and superstitiously thinks that if he carries an umbrella it will rain. We can go on and on forever dreaming up far-fetched combinations of beliefs and desires to attribute to Smith. Any one of these packages of belief and desire will work equally well with [L] to explain why Smith is carrying an umbrella today. To confirm our explanation as the right one, we need evidence that Smith has the wants and desires it attributes to him instead of the far-fetched alternatives.

Suppose we want to be able to predict whether Smith will carry an umbrella tomorrow. To do that, we also need to be able to establish exactly what Smith will want and believe tomorrow morning before he passes the umbrella stand in his front hall. Without this information, we cannot employ [L] to predict what Smith will do. For we can't determine the initial conditions that we need along with [L] to generate a prediction about what Smith will do.

How do we find out exactly what people believe and desire? The most convenient way is of course to ask them. But sometimes people don't want to tell us what they believe and what they want. Instead, we have to just wait till their behavior reveals their beliefs and desires. In some cases we can perform experiments: we can try to arrange their circumstances so their behavior will reveal their beliefs and desires. One thing we cannot do is read their minds. All three methods, asking, experimenting, and observing, are really aspects of the same strategy: they all involve inferring back from action to desire and belief.

Sometimes, in the case of asking, this fact escapes our notice. But a little reflection reveals that asking is just a version of arranging subjects' circumstances and then watching their actions. After all, speech is itself action. Suppose we ask Smith whether he wanted to stay dry today or will want to tomorrow. He emits the noise "yes." Is this an assent to our inquiry? That is, is producing the noise "yes" an action Smith undertook to attain his desire to answer our question in the affirmative, with the belief that producing that noise would do it?

Let's treat the noise that Smith produced as an answer to our question, instead of, say, a funny-sounding sneeze. That involves attributing to Smith at least the following beliefs and desires: (1) the belief that the noise we produced, "Do you want to get wet?" expressed a question in English; (2) the belief that we understand English; (3) the belief that we want an answer to the question; (4) the belief that in English one way to signal assent is to produce the noise "yes"; (5) the desire to signal assent to our question. But even this catalog is not a complete list of desires and beliefs we need to attribute to Smith. Treating the noise he makes as a sign of his desire to stay dry today requires us to add to our assumptions about Smith's desires: (6) the desire to tell us what he believes; and (7) the desire to be sincere and not to lie to us. There are other beliefs and desires we need to attribute to Smith. Identifying them is left as an exercise to the reader, with a warning that the list is too long to complete.

But that means that the easiest way to establish exactly what someone believes and wants is fearfully complex. And what is more, even the sort of first approximation to what people want and believe that we need for everyday life requires us to employ [L] itself. But that raises a brace of serious methodological problems.

The first is a regress problem: to explain an action, we need to identify the beliefs and desires that produced it, in accordance with [L]. To identify those beliefs and desires with any precision, we need to know more about further beliefs and desires. If to do that, we need to know about still further desires and beliefs, the original problem faces us all over again. We have made little progress in answering the challenge to our original explanation. In everyday life we don't face this problem because our explanations are not held to very high standards of accuracy and there is little interest in reducing their vagueness and imprecision. But science requires both challenge and improvement.

Moreover, there is the problem that in order to employ [L] in the explanation of an action, we need to use [L] to establish that the action's causes—the initial conditions—actually occurred. But that means that as long as what is to be explained is an action, nothing could even conceivably lead us to surrender [L] itself. The impossibility of ever being able or willing to surrender [L] casts doubt on its claims to be a causal law. Recall that we want to use Smith's answers to questions as a guide to what he believes and desires. We don't want to treat his "yes" noise as an irrelevant sneeze. Therefore, we have to assume that he wanted to answer our question sincerely and correctly, and that he believed that the way to do so was to use the noise "yes." The reason we have to make these assumptions is that we employ [L] as a guide to understanding Smith's very words. Smith's verbal behavior constituted action, speech, as opposed to "mere" movement of the body, noise, because it was produced in accordance with [L] and Smith's beliefs and desires.

The employment of [L] as a guide to what people believe and desire is even clearer when we have to rely on nonverbal behavior to guess people's beliefs and desires. How can we tell that people believe a ten-dollar bill is worth more than a five, or prefer (desire more) the former to the latter? Offer a choice of one of each to passersby. They all pick the larger bill. But is their behavior a reliable mark of their belief that the ten spot is worth more than the fiver? Only if their behavior reflects their beliefs and desires in accordance with a principle like [L].

In fact, the situation is rather more complicated. To employ behavior as a guide to belief, we have to hold the agent's desires constant. And to use behavior as a guide to his desires, we have to hold his beliefs constant. Any action can be the result of almost any belief, provided the agent has the appropriate desire, and vice versa. Thus, someone might light a cigarette because, say, she believed that the theory of relativity is false. How is this possible? Well, suppose (1) she also believed that someone was asking her whether the theory of relativity was true, and (2) she believed that the way to signal dissent in the language of the questioner was to light up, and (3) she wanted to signal dissent. Bizarre? Well, of course. But that's the point. By itself an action never identifies a single belief or desire. It only does so against the background of a large number of other beliefs and desires.

It's worth emphasizing this point: if we know what someone's beliefs and desires are, then [L] will tell us what actions she will undertake. If we know what actions a person has performed, and we know what her beliefs were, then [L] will tell us what her wants were. And if we know what she wanted and what actions she performed, then [L] will tell us what she believed. But without fixing two of the threesome of belief, desire, and action, the third is not determinable. That is why in explaining action, our aim is to render it intelligible by identifying its meaning or significance. Our aim is to fill in a "hermeneutical circle" of beliefs, desires, and actions, in which coherence among the three variables is the criterion of explanatory adequacy.

It should now be clear why commonsense explanations of human action are so disputable and fallible, and why folk psychology's predictive powers are so difficult to improve. The number of specific beliefs and desires that bring about any particular action is very large. The difficulty of identifying them exactly is great. Accordingly, our explanations of particular actions cannot help being sketchy and lacking in detail. Additionally, they will be subject to considerable doubt, since it is so difficult to nail down much of what a person actually believed and wanted on a given occasion. Our predictions will be no stronger than our explanations. For they rest on nothing but guesswork about the vast number of specific beliefs and desires that are needed for a precise prediction using [L].

Of course, [L] has some predictive content: we can predict with considerable confidence that our professors will not strip off their clothes in the middle of the next lecture; that the driver of the oncoming car will hit the brake as it approaches the stop sign; that a hundred-dollar bill left on the pavement will be picked up; that Dad will have dinner ready when we get home. The reasons for these sorts of predictive success are clear. We can with confidence attribute a certain number of widely held packages of beliefs and desires to everyone, including strangers. And we can attribute more specialized packages to people we know quite well. The number of such predictions we can make with great confidence is indefinitely large. Professors won't strip in class tomorrow, or the next day, or the day after that. Does that mean that [L] has great predictive power?

Predictive power isn't just a matter of numbers of successful predictions. It is based on at least two things: ratio of confirmed to disconfirmed predictions and increasing success in providing highly precise and surprising ones. On both these counts, [L] fares poorly. Beyond obvious predictions, the application of [L] fares poorly. Even knowing a great deal about your family and friends, you cannot predict *exactly* what they will do next; you probably can't even predict *roughly* what they will do next, except in very stereotyped conditions. Strangely enough, most people have been well satisfied with [L] for most of recorded history, despite its weakness. Only twentieth-century social and behavioral scientists have been troubled by their inability to improve [L]'s predictive power. But that's because [L]'s predictive weakness undermines the claim that it is a causal law and thus undermines a scientific approach to human action that employs [L].

It's clear that in order to improve our predictions of human action, we need to do one or both of two things: we need to be able to "measure" people's beliefs and desires with greater precision, and/or we need to improve [L] itself. For example, we need to develop models of rational choice that fill in [L]'s ceteris paribus—other things being equal—clause. That is how all causal explanations and causal laws are improved.

To see that, consider a model from physics: the ideal gas law, PV = nRT. Suppose we want to explain why the pressure gauge on the gas container reads 15.2885 atmospheres. To do so we measure the temperature with a thermometer, discovering it to be 99.5°C, and measure the volume of the container, discovering it to be 2.001 liters. When we plug the temperature and volume into the equation, the result is that the pressure is 15.3101. This value is probably close enough to the true value for most purposes. Indeed, all measuring instruments, thermometers, manometers, and meter sticks, have margins for error. For all we know, the theoretically derived value for pressure may be closer to the real value than the pressure gauge's reading. How can we decide which is more accurate—the value predicted by plugging the data into the equation or the value observed on the pressure gauge? The only way to decide is by improving our measuring instruments. Substitute a thermometer that reads out digitally to more decimal places; a micrometer instead of a meter stick to measure length, breadth, depth, for volume; and a digital pressure gauge instead of an analog dial. Any of these changes will help give more accurate measurements of the initial conditions. If the more accurate numbers provided by these new measuring instruments are plugged into the equation, the resulting calculation may be much closer to the observed values. In other words, the precision of our explanation will have been improved, and the accuracy of future predictions using the ideal gas law will also have been increased by these new instruments.

Even more important is what happens when improvements in measurement of initial conditions do not result in calculated values closer to observed ones. Such an outcome disconfirms the law. When improved measurements do not vindicate the equation into which they plug data, we try to improve the equation, model, or generalization. In fact, the history of improvements and complications in the kinetic theory of gases involves successive refinement of the equation to bring its predictions into line with the data provided by better measuring instruments on more and different gases. When new variables and constants are added, the theory can explain and predict the behavior of a gas to a greater degree of accuracy, over a wider range of values of pressure, temperature, and volume. These improvements rest on our ability to measure initial conditions with increasing accuracy over wider and wider ranges.

But neither of these sorts of improvements is possible for [L] or for the explanations and predictions in which it figures. PV = nRT can be applied to explain and predict in part because there are thermometers to measure initial conditions. What functions as the "thermometer" for [L], the means for measuring its initial conditions? In too many cases, the only measuring instrument available is [L] itself. In almost all cases, [L] must serve as its own "thermometer": To measure people's beliefs and desires we have to use [L]. Moreover, although we use different instruments to measure temperature, pressure, and volume, we need the same instrument—[L]—to measure all three of our explanatory and predictive variables.

But that means that we start out with a "law," [L], of limited predictive power, given the difficulty of establishing its initial conditions of application with any precision. Then there is little chance to improve it. For to improve it, we need first to find cases where [L] has gone wrong in its prediction; second, to "measure" the values of the initial conditions and the actual behavior that it failed to predict correctly; and third, to revise our "measurement" to accommodate the observed action. But in order to "measure" beliefs and desires, we must use [L] itself, plus the observed action we failed to predict, and then work back to a more accurate determination of the beliefs and desires. Once we've done that and plugged the more accurate initial conditions into the predictive argument, [L] gives us the observed action. But then there is nothing to correct about [L] after all. There is never an opportunity to add to or subtract from [L] in order to improve it.

One popular way of describing this problem for [L] is to say that it is unfalsifiable: there is no conceivable evidence about human action that could lead us to surrender it.

We can imagine what it would be like to falsify a scientific model, generalization, or law. Imagine jumping off a cliff and just hanging in midair instead of falling with constant acceleration. That would falsify Newton's first law. The demand that to be a law or to describe any causal relationship, a statement must be falsifiable is probably the most held methodological principle in the social sciences. Everyone recognizes that a law can never be fully confirmed or verified since it is about an indefinite number of events in the future, the present, and the past. But apparently it takes only one wrong prediction to falsify a proposed law.

The methodological principle is that to be a real law with explanatory power, a generalization must be one we could imagine or conceive to come out false under some circumstances. Notice two things: First, the requirement is not that to be a law a statement has to actually be falsified; no, the test is whether we can describe circumstances that would show it to be false, not that we actually find such circumstances. Second, notice that the truths of mathematics, though completely general, are not falsifiable. We cannot imagine circumstances under which $4^2 = 16$ is false. That is why mathematical truths can't by themselves explain particular facts in the world. The same goes for definitions: "All bachelors are unmarried" won't explain why some single male friend of yours is a bachelor. And part of the reason presumably is that "All bachelors are unmarried males" has no empirical content; it's a definition that no possible factual evidence could overthrow. It's unfalsifiable.

But it is hard to think what could falsify [L]. Remember, irrational behavior isn't action at all. But if [L] is unfalsifiable, then it cannot provide empirical, scientific knowledge. Therefore, if it does provide knowledge essential to social science, then social science's explanations of human action that use [L] are not ultimately empirical ones. Rather, they are interpretative. They help us see the meaning of the actions [L] explains. What is more, social science's failure to provide predictions beyond those of folk psychology reflects no discredit upon it. The lack of predictions merely reflects the differences in its aim from that of the natural sciences. That of course is the interpretationalist's or antinaturalist's view of [L].

Introduction to the Literature

The centrality of action explanations to social science is emphasized in both the methodological and the substantive work of Max Weber, one of the founding fathers of sociology. The debate about what Weber called *Verstehen* (empathetic understanding) and "ideal types"—unrealistic assumptions in explanatory models—began before his work, but its twentieth-century form is a result largely of his formulation of the issues. See Weber, *The Methodology of the Social Sciences*. The role and nature of idealizations and models in social science are treated in his Chapters 3 and 6.

Several papers in Martin and McIntyre's *Readings in the Philosophy of Social Science* are directly relevant to the question whether there are or can be laws of any kind in the social sciences. Of particular value are the papers by B. Fay, "General Laws and Explaining Human Behavior," which argues for a causal account of reasons without laws; and H. Kincaid, "Defending Laws in the Social Sciences," and McIntyre, "Complexity and Social Scientific Laws," which defend the possibility of naturalism about reasons and actions. Papers in Martin and McIntyre focusing on whether folk psychology can provide laws include D. Davidson, "Psychology and Philosophy," and W. Dray, "The Rationale of Actions." D. Henderson, *Interpretation and Explanation in the Human Sciences*, is an extended attempt to reconcile the holism of the mental with naturalism and to defend the lawlike status of [L]. Paul Churchland's paper "The Logical Character of Action Explanations" defends [L] as a lawlike statement cogently. Papers by Hempel and Kincaid on this subject are also reprinted in Steel and Guala, *The Philosophy of Social Science Reader*.

The demand that there be a law connecting desires, beliefs, and actions reflects the thesis that scientific explanation must involve laws. For a long time this question was the most widely discussed topic in the philosophy of social science. Several of the most influential arguments in favor of the covering-law model's applicability to human action were written by Carl Hempel and have been collected in an anthology of his papers, *Aspects of Scientific Explanation*. See especially "The Function of General Laws in History." See also his essay, "Rational Action," in *Proceedings of the APA*, 1962. These views are challenged in William Dray, *Law and Explanation in History.* Dray's *Philosophy of History* discusses this issue lucidly. Karsten Stueber updates this debate in "The Psychological Basis of Historical Explanation: Reenactment, Simulation, and the Fusion of Horizons" in Steel and Guala's anthology, *The Philosophy of Social Science Reader.*

For a contemporary discussion of alternative approaches to the understanding of belief/desire explanations as they are involved in interpretation, see Alvin Goldman, "Interpretation Psychologized" in Steel and Guala's anthology.

CHAPTER FOUR

Intentionality and Intensionality

Any attempt to reconcile a causal treatment of the explanation of action in terms of desires and beliefs faces deep problems that require a certain amount of complex philosophical apparatus. This chapter introduces and employs some much-needed technical vocabulary that will recur later in the book. Using this philosophical equipment enables us to trace the problems facing the social sciences regarding the question of how the mind is related to the body, a problem with a long pedigree in philosophy.

THE LOGICAL CONNECTION ARGUMENT

In the roughly two decades before 1970, arguments like the one at the end of Chapter 3 provided the main challenge to the notion that social science is a form of causal inquiry. The arguments were simpler than that just given and were unsound, but they reflected a partial realization of the complex problems facing a causal approach to the explanation of action. In fact, the more complex argument just given is the result of reflections on the simpler ones. The simpler arguments began with the assertion that causal claims had to be contingent truths, not necessary truths or definitions, and that causal explanations required contingent generalizations. Thus, for example, no one could causally explain why Smith is a bachelor by pointing out that he is an unmarried man and all unmarried men are bachelors. Being a bachelor and being an unmarried male are logically connected. It is inconceivable that a bachelor not be an unmarried male. An instance of the first is identical to some instance of the second. The generalization here is a definition, and the initial conditions in effect redescribe the fact to be explained. But just as nothing can be its own cause, nothing can causally explain itself. Thus, whatever enlightenment is provided by the information that Smith is unmarried, it cannot be part of a causal explanation of why Smith is a bachelor. If [L] is not a causal law about action, how does it function in explanations?

Every explanation of a human action is in fact tantamount to a redescription of the event to be explained. It does not identify other distinct and logically independent events, states, or conditions that determine the action. If a statement like [L] is essential in the explanation of human action, it is because [L] is part of what we need to make the redescribing explanations. [L] defines what it is to be an action in terms of the notions of desire, belief. [L]'s function is to show us what counts as having a reason for doing something and to show us when a movement of the body is an action. Thus, desires, beliefs, and actions are logically connected, not contingently connected, by [L]. Therefore desires, beliefs, and actions are not causally connected by [L] or by any single causal law.

The role [L] plays in identifying beliefs and desires seems to testify to this claim. So does the fact that no failures to predict human action would lead us to give up [L] or to identify some improvement on it. It is pretty evident that [L] is not the result of any self-conscious experimentation, observation, or data collection a social scientist undertook with a view to framing a general law about human behavior. It is, as we have noted, a piece of folk psychology, embedded in our consciousness as far back as we know. [L] is a principle that reflects our conception of ourselves as responsible agents, as persons who act on reasons. If nothing could lead us to give it up, then the connection [L] expresses between desires, beliefs, and action cannot be the sort of causal link reported in our scientific laws. For science is fallible, and any of its generalizations can be given up.

Recall the brief account of causation in Chapter 2 ("Progress and Prediction"). Its claim was that causal links must be contingent. That is, when event *a* causes event *b*, it must be conceivable for either to have occurred without the other's having done so. That is why being an unmarried male will not explain why someone is a bachelor. It is inconceivable for someone to be one and not the other. There is a logical connection between them. If that were not the case, then knowledge of the occurrence of the effect would be enough to establish the occurrence of the cause, and vice versa. But in science, knowledge about any single event never is enough to establish the character or occurrence of another event.

To determine something's effects or causes, we need to do empirical research, undertake experiments, conduct observations, collect data. This empirical evidence is expressed in contingent laws and generalizations, propositions that could conceivably be false, unlike definitions. These laws sustain our singular causal claims, even when they are unmentioned in the claims. Consider the true statement that striking an iceberg caused the *Ti-*

tanic to sink. This claim about two particular, distinct events, the striking and the sinking, is justified by a large number of physical laws. Of course none of these laws mentions the *Titanic* or ships of any kind. They are laws of buoyancy, principles about the tensile strength of materials like cast iron and ice, and so forth. Most people who correctly claim that the *Titanic*'s striking the iceberg caused its sinking don't know any of the physical laws. Nevertheless, making the claim that it was the striking of the iceberg that caused the *Titanic* to sink commits those people at least to the existence of such laws, even if they don't know what the laws state.

Now, the argument continues, return to the explanation of actions by appeal to beliefs and desires. It's clear that our description of Smith's belief that carrying an umbrella today will help keep him dry is logically connected to Smith's carrying the umbrella, the alleged effect of that belief. For the description of the belief makes reference to the action itself. We could not characterize or describe this belief except in a way that refers to one or another of its alleged effects. And we cannot describe what a person does as an action without thereby committing ourselves to the existence of desires and beliefs that contain a description of what the person does as an action. But, the logical connection argument holds, this commitment precludes the existence of contingent connections between desires, beliefs, and actions. For they can be described only in terms that refer to one another. Accordingly, such connections will not be causal, and any generalizations that link them will not be general laws either.

Rather, [L], statements like it, and their consequences will be necessarily true definitions and their consequences that provide some form of noncausal explanation. That is why [L] is not open to rejection or improvement as a result of empirical findings about contingent matters of fact. Philosophical analysis of [L] thus reveals it to be a definition, one useful for interpreting action and rendering it intelligible. It is a philosophical confusion to view it as a contingent generalization, a mistake that leads to the sterility and frustration characteristic of naturalistic social science.

However, this argument is too strong. It doesn't really prove that beliefs and desires can't be causes of action. Nor does it establish that statements like [L] can't be treated as causal generalizations. But it does reflect a real difficulty for the attempt to treat [L] and its explanatory applications in the way the naturalist wants to: as laws or precursors to them.

To see what is wrong with the logical connection argument, we have to distinguish events and processes from the descriptions we give of them. An event, a thing, or a person can be described in many ways: George Washington may be described as the father of his country, the first president of the United States, the owner of Mount Vernon in 1799, Martha Washington's husband. Sometimes the description of one thing includes reference to another. Thus, describing George Washington as "Martha's husband" makes reference to Martha as well as to George. Sometimes a description may refer to a future time and place, as when we say, "The man who was to become the first president of the United States was born in 1732."

Similarly, a cause and its effect can be described in many ways. The sinking of the *Titanic* can be described both as "the disappearance beneath the waves of the pride of the White Star Line" and as "the greatest loss of life at sea in 1912." Thus, we say the same thing when we say, "the *Titanic*'s striking the iceberg caused it to sink," and when we say, "the striking of an iceberg by the fastest ship on the Atlantic in 1912 caused the sinking of the largest vessel in the White Star Line." Someone ignorant of the fact that the fastest ship on the Atlantic in 1912 was also the largest ship in the White Star Line might find the second causal claim less informative than the first. Some people might even believe the first claim true while denying the second. Most of us would consider the second claim misleading because it seems to suggest that two ships were involved, the fastest one and the largest one. But both statements are true, and both report the same causal sequence.

Let's consider the following claim: The sinking of the *Titanic* was caused by the event that caused the *Titanic* to sink. Well, that looks like the degenerate case of an uninformative claim. Is it still a causal statement, one asserting a contingent connection between two distinct events? Or is it true by the definition of the word *cause*? Once we bear in mind that the same event may be described in many ways, our answer to this question must be that it still reports a contingent causal claim. For the *Titanic*'s striking the iceberg is identical to the event we describe as "the cause of the *Titanic*'s sinking." If we put the former description into the sentence above, we get the right answer: the sinking was caused by the striking.

But in our degenerate case, the events are *logically connected* in that the description of the cause refers to the effect. Yet the connection it claims to obtain is still causal. True, it is no longer explanatory; it is, as we said, uninformative, useless in any significant inquiry; it tells us almost nothing more than we already knew. But it does tell us something nontrivial, nondefinitional, namely, that the event, the *Titanic*'s sinking, occurred and that it had a cause.

Suppose, however, that we didn't know any more about the *Titanic* than that it sank and that this event had some cause or another. This is the situation we would be in if we knew nothing else about the event that caused the sinking besides the uninformative one that it caused the sinking. We would be in the position of insurance investigators or detectives who are confronted with, say, a fire and must find an explanation for it. They search for another true and informative description of the event uninformatively described as “the cause of the fire.” Causal inquiry always starts with this sort of ignorance, reflected in an uninformative description of the cause logically connected to the effect. But the investigator’s inquiry is still a causal inquiry. It aims at identifying another distinct event by providing at least one description that is not logically connected to the description of the fire. So, just because some true descriptions link two or more distinct events logically, it does not follow that all descriptions of the same events do. Nor does it follow that where logical links between the descriptions of events obtain, there can be no causal links between the events they describe.

The moral of the story for beliefs, desires, and actions should be clear. Beliefs and desires are usually described in terms that make reference to the actions they produce. So there is a logical link between the descriptions we actually use to identify beliefs and desires, on the one hand, and descriptions of actions, on the other. But from the existence of the logical link, it does not follow that beliefs and desires are not causes of actions or that [L], some improvement on it, or perhaps an altogether different generalization about actions and reasons cannot be laws of human action.

Thus, the logical connection argument turns out to be too weak to show that beliefs and desires cannot be causes. But it does reveal something very important about explanations of actions by beliefs and desires. It reveals why they often seem so empty. If I say, “Smith carried an umbrella because he wanted to,” that may be a causal claim, but it is as close to empty a description of the cause of Smith’s action as one can come without getting there completely. It’s like saying, “The *Titanic* sank because it was caused to sink.” Saying, “Smith carried an umbrella because he believed doing so would help attain his desires,” is only slightly more informative. The explanation employs information that would provide a means to predict what Smith will do the next time he has a chance to carry an umbrella.

Beliefs, desires, and actions are distinct, causally connected states and events. But what if we can never actually discover any descriptions of beliefs, desires, or actions that link them contingently in laws instead of logically in definitions? Then we will be in the position of detectives who know in principle that their arson case won’t be solved, because they can’t uncover any informative description of the fire’s cause. All they can ever know is that the fire had a cause, because fires don’t start spontaneously.

This conclusion, that there are no descriptions of beliefs, desires, or actions that can even conceivably be independent of one another, is one that the philosophy of psychology has had to take very seriously. In the next section we will see why independent descriptions of reasons and actions are hard to uncover. The ramifications of this fact for the social sciences are very great. For if this conclusion is right, even though desire/belief explanations of actions do link causes to their effects, there are no general models, regularities, or laws we can discover to link them together. But in the absence of such laws, the explanatory power of reasons—desires and beliefs—cannot in the end be causal. That suggests to some social scientists, especially behaviorists, that we should surrender the goal of explaining actions by uncovering desire/belief combinations. Instead, they hold, we should search for more satisfactory causal explanations of human behavior, ones that appeal to laws we can discover.

Among interpretationalists, this same failure to uncover laws about reasons is an argument for surrendering the aim of finding causal explanations of action. Instead, we should attempt to determine in what the evident and unarguable explanatory power of folk psychology consists. This is the strategy of interpretative social science. Chapters 7 and 8 take up these strategies in turn.

INTENTIONALITY

Why suppose that we can never find descriptions of actions and the reasons that cause them that are independent enough from one another to enable us to frame laws about them? All three of the variables of folk psychology, desire, belief, and action, are *intentional.* This term has a special meaning in philosophy, though one related to its ordinary meaning of "purposefulness." To say that a belief is intentional in the philosopher's sense is to say that it has "propositional content," that beliefs "contain"—in some sense—propositions or statements or sentences. Thus, there cannot be a belief without a proposition believed. Belief, it is often said, is a relation between a sentient creature and a statement: *x* believes that *p*. When we state one of our own beliefs or claim another has a belief, there is linguistic expression following the "that" in our statement. That expression is a grammatically complete sentence, one that expresses some proposition—true or false. Thus, "Stalin believed that Hitler would not attack Russia" is a sentence that contains another sentence within it: "Hitler would not attack Russia." The belief that the whole sentence attributes to Stalin presumably must itself "contain" this statement. That desires are intentional is less obvious. But that is because we frequently abbreviate the whole sentence and the "that" that desires contain. Thus, when I say, "I want an ice cream cone," what I desire is that it be the case that I have an ice cream cone—notice the "that" clause and the whole sentence. English allows us to abbreviate the "that" clause into its object-noun. Or, to give another example, my desire to go to the movies is the desire that I go to the movies. Here the content of my desire is again a whole grammatical sentence that expresses a complete proposition. English grammar obscures this fact by dropping the "that" and changing the present indicative form of the verb "I go" to the infinitive "to go": "I want to go to the movies." Desires, beliefs, hopes, fears, almost all cognitive states, are intentional: they have "propositional content"; they are "psychological attitudes" toward statements.

The quotation marks around the words *contain* and *content* remind us that no one understands exactly how the physical matter in our brains can "contain"—store and retrieve—statements. The use of the term *contain* may turn out to be at best metaphorical. The puzzle about how the gray matter of the brain can represent the way the world is or, in the case of a desire, the way someone wants it to be has been a central problem for philosophy since its beginnings. It is not an easy problem to recognize. But many of those who oppose a naturalistic approach in the behavioral and social sciences base their opposition on the alleged insolubility of the problem of how the brain represents the world.

When we say that brain states contain statements about, or representations of, the world, the term *contain* is metaphorical because it is hard to take seriously the suggestion that the statement is "written" in any language on or in the synapses of the brain. Why? Because to take this claim literally seems to involve an absurdity. Recall library card catalogs, before the days of computerized catalogs. Each of the ink-marked cards in the catalog represented a book. But they did so because there were library users who interpreted the ink marks. The ink marks and the cards they were on didn't intrinsically and directly represent anything; they were just pigment on pieces of wood pulp. It's perfectly conceivable for such ink marks to have been formed on pieces of wood pulp by accident, in the way that a cloud might resemble a face, or a tree might naturally take on the shape of the letter *Y*, without the intervention of an agent who wanted it to represent the letter *Y*. Only the existence of creatures who treated the ink marks as having a meaning could give the cards their representational character.

Now consider the gray matter of the brain. When a statement about the world, like "the sun is setting," is represented in one part of the brain, *b*', who is the interpreter who treats the configuration of synapses at *b*' as expressing this statement? Can we say that the mind reads the meaning off the configuration at *b*'? Yes, but then if we think the mind is distinct from the brain, we face the problems of how a nonphysical mind can represent and how it can possibly influence physical matter like the brain.

Is it more attractive to suppose that there is some other part of the brain, *b*", "reading" the statement from the gray matter at *b*'? If we are tempted to say yes, we must face the same question all over again for *b*”, the part of the brain that does the reading. For it to read the statement from *b*’, it must represent, somewhere within it, the meaning of what it reads in the synapses at *b*’. Let’s call the part of *b*” that represents the statement that *b*” reads from *b*’ the subsystem *b*’”. But the same question arises for *b*’”. We are on an infinite regress, and we have explained nothing about how physical matter can represent the world to itself.

Because of this problem, some philosophers and psychologists hold that the representational powers of human thought will never be explained in terms of neural processes in the brain or by any other purely physical system. They argue forcefully that psychology and the social sciences cannot be naturalistic enterprises. For these disciplines deal in beliefs and desires, the very vehicles of representation that cannot be explained in a scientific way.

The philosophy of psychology and many psychologists grapple with the mystery of how we represent the world to ourselves while viewing ourselves as purely physical systems. Until they solve this puzzle, we will not understand how beliefs and desires “contain” statements. But we will continue to identify and distinguish beliefs and desires from one another on the basis of their propositional content. Thus, the difference between your belief that Paris is the capital of France and your belief that Warsaw is the capital of Poland is given by these two different sentences that your beliefs “contain.” And presumably your belief that Paris is the capital of France is an instance of the same type as my belief that Paris is the capital of France because the statements they contain are identical.

One indication that desires and beliefs “contain” statements is grammatical: the sentences that describe them literally contain sentences, introduced by “that” clauses. But being described by sentences containing “that” clauses is not what makes beliefs and desires intentional. Lots of sentences in which “that” clauses figure are not intentional: “The *Titanic*’s striking the iceberg caused it to be the case that the *Titanic* sank.” This is a nonintentional sentence containing a “that” clause introducing another whole sentence: “The *Titanic* sank.” What shows that a belief or a desire is intentional is not the grammar of the sentence describing it. The intentionality of psychological attitudes is revealed when certain apparently innocent substitutions are made in the sentences that describe them.

This is easier to explain by illustration. Take the sentence “Lois Lane believes that Superman is courageous.” Now, there are many things that Lois Lane does not know about Superman. In particular, she does not know that Superman is identical to Clark Kent. But he is. Now suppose we substitute “Clark Kent” for “Superman” in our statement of Lois Lane’s belief. This substitution will turn our description of Lois Lane’s belief from a truth to a falsity, for she does not believe that Clark Kent is courageous. Indeed, she believes the contrary. By contrast, make the substitution of "Clark Kent" for "Superman" in the statement "Superman is courageous" and the result will remain true: Clark Kent is courageous; it's just that few people know it. We might be tempted to say that Lois Lane does believe that Clark Kent is courageous, but not *under that description*. Our example shows that the terms used to express a belief are crucial to correctly identifying it, in a way that they are not crucial to identifying nonintentional facts. An intentional description is one that can be changed from true to false just by substituting nouns, adjectives, verbs, and adverbs that refer to the same things. An intentional state—like a belief—is one whose description is intentional.

This sensitivity of intentional states to the descriptions and terms we use to identify them is even clearer in the case of desires. Thus, "Lois Lane wants to marry Superman" is a true statement about that enterprising reporter's desire. But Superman is identical to Clark Kent. And under that description of Superman, Lois Lane wants no part of him. If we substitute "equals for equals" in the true statement that she wants to marry Superman, we get the false one, "Lois Lane wants to marry Clark Kent." Lois Lane wants to marry Superman "under one description," but under another she does not want to do so.

Thus, intentional states are ones of which we cannot freely substitute synonymous descriptions without risking the chance of changing a truth to a falsity. That should not really be surprising. Beliefs and desires are "subjective." They reflect the ways we look at the world, our points of view, which differ from one another and change as our perspectives change, as we acquire different information about the world. A representation of how things are or could be must always be drawn from a perspective, one that is partial and incomplete. The subjectivity of our beliefs about the world is reflected in this curious feature—that substitutions that make no difference to truths about objective states of the world make a great deal of difference in descriptions of subjective states.

Philosophers have a special term to indicate the sensitivity of intentional statements to substitutions. They call those statements "intensional"—with an *s* instead of a *t*. The terminology is regrettable because it breeds confusion between intentionality, a property of psychological states, and intensionality, a logical property of statements that report psychological states. One reason philosophers use this new term *intensionality* is that there are other statements besides those reporting intentional, that is, psychological, states; those other statements are intensional, that is, sensitive to equals-for-equals substitutions in their terms. We shall not concern ourselves with these other statements.

The intentionality of desires and beliefs makes the explanation of actions intentional as well and thus makes actions derivatively intentional. Consider the Spanish explorer Ponce de León, who searched for the Fountain of Youth. In this case, searching cannot have been a relation between Ponce de León and any particular object hidden in the uncharted jungles of southern Florida. For there is no such object. Yet the description "searching for the Fountain of Youth" must be related to Ponce de León's behavior, for that is what he was doing. What makes his behavior the action of searching for the Fountain of Youth is that it was produced by beliefs and desires that contain statements about the Fountain of Youth. So the explanation of action is intentional because it results from intentional states like desire and belief.

Intentionality turns "mere" behavior into action. Action is intentional, for behavior is only action if there are intentional states—desire and belief that lead to it. Since desires and beliefs "contain propositions," their effects—actions—reflect the propositions they contain as well. Thus all the apparatus that common sense and social science employ to describe what people do (as opposed to what merely happens to them) has an intimate connection with language. For it is the sentences of a language that give the content of desires, beliefs, that express the relation between us and propositions about the world. It's not just that these states "contain" statements. The statements psychological attitudes contain give them their identity. What distinguishes one belief or desire from another is the difference in the sentences each contains. When two different people have the same desire or belief, the sameness is due to the identity of the statement they believe or want to be true. Change the terms in which a statement someone believes is expressed, and you may well change the belief itself.

To explain an action with full precision, one must identify the very sentence in which the agent would describe his action and therefore the very sentences that the agent would use to describe his beliefs and desires. The fact that we can rarely if ever do that certainly puts limits on the precision with which we explain actions after the fact and restricts even more our powers to predict actions as yet not taken.

It will come as no surprise to many social scientists that language is intimately connected with action and its explanation. This account of intentionality may be the strongest argument for such a claim. Many social scientists and philosophers have long held that the aim of social science is to reveal the meaning of behavior, or its significance. And they have usually contrasted meaning and significance to causation as incompatible alternative aims. The analysis of the description and explanation of action as intentional doesn't just give new force to this idea. It provides a concrete argument for the indispensability of meaning to action. To give the meaning of an action is now taken out of the realm of the metaphorical and made an essential step in explaining it. We cannot explain an action till we know what action it is. We cannot know what action it is unless we know how the agent views it, that is, under which linguistic description the agent brings it. Once we know that, we can explain the action by showing its significance, its role in meeting the agent's desires, given his beliefs. Since both desires and beliefs are meaningful states of an agent, the explanation they provide action gives its meaning in a very literal sense.

Some philosophers, and many cultural anthropologists, have likened social science to language learning. Others have held the stronger view that social inquiry is nothing but language learning. Thus, once anthropologists had gone native and learned the language of a hitherto strange culture, they acquired all the resources needed to explain the actions of their subjects. The anthropologists have uncovered as much theory as there is to uncover, for now they know the terms in which their subjects describe behavior and express to themselves the beliefs and desires that produce it. This view of social science is pretty clearly an interpretationalist one, and we shall return to it in Chapter 8. For the moment, it is enough to see how much this view is fostered by the intentionality of human action and what problems it makes for naturalism.

INTENSIONALITY AND EXTENSIONALITY

The immediate upshot of the intensionality of action and its determinants is that it seems to make the causal approach to human action impossible. For it shows that there is no way, even in principle, of providing a description of the beliefs and desires that cause action in which they are independent of one another and independent of the action they are said to cause.

Recall the admission that [L] is employed in everyday life and in social science to establish initial conditions, which are then harnessed together with [L] again to explain an action. Indeed, [L] is also employed to determine whether a bit of behavior is action. For [L] not only links desires and beliefs to the action they explain; it is also our tool for identifying what beliefs and desires the agent has. This multiple use of [L] to carve out its own domain of explanation, as well as to establish its own initial conditions, is not logically illegitimate. But it is methodologically suspect, for it makes it difficult to surrender [L] in the light of any possible empirical evidence. That means [L] lacks much empirical content. Among social scientists who demand that their theories have substantial testable content, [L] will not fare well. [L] certainly does not seem to be falsifiable by any conceivable observable evidence: whenever a person does something that looks utterly irrational, given the beliefs and desires we have attributed to him, the reasonable thing to do is change our estimate of his beliefs and desires; in the light of really crazy beliefs and/or desires, any action will look rational. We could, of course, decide that the behavior really is irrational, but then we wouldn't explain it as action, but rather as a form of pathological behavior. What we cannot do is give up [L] to preserve the original estimates of an agent's beliefs and desires. For [L] is what we use to estimate the beliefs and desires we would be giving up [L] to preserve. If our explanatory generalizations must be falsifiable, then [L] must be surrendered as a causal law.

But the requirement that our explanatory hypotheses be falsifiable is generally viewed as too strong a demand on scientific theorizing. It neglects the fact that scientific hypotheses do not make claims about observations by themselves. To test a hypothesis, we need to add initial conditions, just as we need to add initial conditions to a law when we explain an event. But this actual or merely conceivable falsification would cast doubt on a set of statements, including the hypothesis being tested and the assumptions we used to establish or measure the initial conditions. Under these circumstances, a favored hypothesis can be retained, no matter what observations are made, by making suitable adjustments in the assumptions we must make in order to test it.

For example, as we saw in Chapter 3, PV = nRT cannot be tested by observing the temperature, pressure, and volume of a gas unless we measure the gas's temperature. But using a thermometer is adopting a hypothesis about how heat affects the length of a column of mercury, because that is the hypothesis that guides thermometer construction. Any divergence between observed and predicted values for pressure, given temperature and volume, can always be blamed on faulty measuring devices or on the falsity of the hypothesis about thermal expansion of the mercury columns. After all, if we use water instead of mercury in our thermometer, we will not find any gas whose temperature will exceed the boiling point of water. Without realizing that our hypothesis about a thermometer using water is false, we might conclude that PV = nRT is falsified. Clearly, in this case we must revise our measuring hypothesis instead of the ideal gas law.

Because of the role of auxiliary hypotheses, no general law is strictly falsifiable. Therefore, [L] can hardly be held to this standard. The real problem for [L] isn't testability; it's that in applying and improving [L], we need to formulate the right sort of auxiliary hypotheses about desires, beliefs, and actions, ones that will enable us to test and improve it. Even in the absence of such auxiliaries, [L] conveys some minimum causal information. Its empirical content is illustrated in its powers to guide our most basic everyday expectations about how others will and will not behave. But if we could provide an alternative means to establish [L]'s domain and its initial conditions, then we could in fact proceed to test [L] and begin to improve it.

Now, what the intensionality of our descriptions of beliefs, desires, and actions suggests is that no such alternative means will ever be found. There are only two sources for a determination of what a person believes or desires: behavior and brain states. What we need is something that will "measure" what a person believes by some distinct effect of the belief, in the way that a thermometer measures heat by its quite distinct effect, the height of a column of mercury or alcohol. We need an equation with a belief (or a desire) on one side of the equal sign and a brain state or description of behavior on the other. Some neuroscientists and others, including participants in the growing field of neuroeconomics, believe that we are on the way to providing ourselves with such identities. Functioning magnetic resonance imaging, fMRI, already enables us to localize mental processes to distinct brain regions. Why should it or some successor not eventually enable us to read particular thoughts "off" the brain by imaging techniques?

Skeptics about the powers of neuroscience or behavioral science to do this will argue that such an equation is impossible because something will always be missing from the brain or behavior side of the equation: intensionality. The description of behavior or brain states is never intensional. It is, in the philosopher's lingo, "extensional." In fact, all the rest of science—biological, chemical, physical, and mathematical—can be expressed in extensional terms. Consider any true description of a bit of mere behavior or of a brain state, whether in the language of anatomical size, shape, location, physiology, cytology, molecular biology, chemistry, or electromagnetic theory. The description will remain true even when we substitute equivalent descriptions into it, no matter how far-fetched.

Providing an equation that identifies the extensional facts about brain states or behavior in which a belief or desire consists in fact constitutes a solution to the mind-body problem. Of course, we can just assert that every intentional state is identical to some brain state or other. But that is no solution to our problem, for it does not enable us to identify the belief or desire that any particular brain state constitutes.

Recall the point from Chapter 3. If what I believe is a function of all or most of my other beliefs and desires, then my belief state is identical with the state of my whole cerebrum. But no description of that much of my brain could be used to provide a useful "thermometer" to measure a single psychological attitude toward a particular proposition.

If the intensionality of mental-state descriptions is missing from the description of the brain states we attempt to use as indicators, then the equation must be wrong. In contrast, if the description of brain states or behavior is intensional, we have not solved our problem but simply shifted it to a new level. For now we will need a test for the intensional content of a piece of behavior or a brain state. And no amount of extensional neuroscience can give us such descriptions.

Philosophers of psychology have expressed this point by saying that mental states are not reducible to behavior or brain states. The problem they face is the ancient one of the mind and the body: How is the former related to the latter, and what kind of a thing is the mind anyway? Philosophers of the other social sciences may think they can ignore such arcane questions. But in the end they cannot. For the modern version of the mind-body problem is that of how physical matter can have content (can "represent") in light of the fact that a complete description of it will be extensional and never intensional. It becomes a problem for the philosopher of social science when the role of folk psychology in the explanatory strategy of social science is made clear. For the only way to improve on folk psychology's unity and precision is by showing the "measurability" of its causes by means that the rest of science recognizes. Only if such linkage is possible will there be, even in principle, alternative means for identifying [L]'s domain of application. Only if such linkage is possible will there be means independent of [L] to determine the occurrence of its initial conditions. Since such linkage is impossible, it looks as if the conclusion of the logical connection argument is right after all, even though the argument is unsound. For there is no description, known or unknown, of the intentional causes of action, no description that is itself extensional and thus none that is independent of a description of their effects. [L] therefore turns out not even to be of limited use as a causal regularity, for the elements it connects cannot, even in principle, be shown to bear contingent relations to one another.

This is a pessimistic conclusion for the naturalist who hopes to meet scientific standards in the explanation of human action. Whether it is too pessimistic hinges on the resolution of fundamental metaphysical problems about the nature of the mind and its relation to the body. The pessimistic conclusion rests equally on deep epistemological issues about the possibility of empirical testing and its relation to scientific knowledge.

Some naturalists believe that [L] and intentional folk psychology generally can still be reconciled with a broadly empiricist approach to psychological theory. They accept the fact that folk psychology will always be with us; no one can give it up. Indeed, no one should give it up. To begin with, for all its predictive weakness, there is no other theory of human behavior available that is predictively more powerful; second, folk psychology has in fact been improved in the way required for science in at least some areas of cognitive psychology. Accordingly, those naturalists argue, our conception of the nature of scientific theories needs to be changed, improved, enriched, until it can accommodate statements like [L] as lawlike generalizations. They argue that we need to rethink our conception of science until [L] and its like are rightly viewed as laws, despite their predictive weakness and the difficulty of linking them up to models, generalizations, and laws in nonintentional psychology and the rest of science. The largely philosophical debate that this approach has generated cannot alleviate the pessimism about folk psychology that has characterized much empirical social science in the second half of the twentieth century.

In Chapters 5 and 6 we will explore the effects of this sort of pessimism on some naturalistic social scientists. In Chapters 7 and 8 we turn to the ramifications of this pessimism about naturalistic theories of human action for the opposing view. We shall see how antinaturalists draw optimistic conclusions from it about the character of a nonscientific approach to human behavior.

Introduction to the Literature

The claim that [L] and propositions like it are necessary truths is defended in several works, including P. Winch's *The Idea of a Social Science*, R. S. Peters's *Concept of Motivation,* A. Melden's *Free Action*, and most ably in Charles Taylor's *Explanation of Behavior*. These works are the source of what has come to be called the logical connection argument. The most powerful rejoinder to these arguments is to be found in D. Davidson, "Actions, Reasons, and Causes," in *Essays on Actions and Events.* More recent discussion of [L] as an interpretative principle is to be found in papers by Taylor, "Interpretation and the Sciences of Man," and Michael Martin, "Taylor on Interpretation and the Sciences of Man," in Martin and McIntyre. The logical connection argument is no longer widely advanced in the philosophy of social science, and to that extent the discussion of this chapter has the character of an exposition of a bit of the twentieth-century history of the subject. The eclipse of the argument is due largely to developments in the philosophy of psychology, including especially the work of Dennett and Fodor. Most of the debate about whether reasons can be causes, nomological or otherwise, has shifted to the question of how purely mathematical models of rational choice characteristic of economics can explain. This is the subject of parts of Chapter 3 of this book.

Problems of intentionality and intensionality are among the most vexed in philosophy. An excellent introduction to the contemporary philosophy of psychology and to the problems of intentionality is Paul Churchland, *Matter and Consciousness*. Problems facing the causal analysis of the mind have been expounded by W. V. O. Quine, Donald Davidson, Daniel Dennett, and others. Important papers by these figures are anthologized in N. Block, ed., *Readings in the Philosophy of Psychology*. The contemporary locus classicus of the debate is D. C. Dennett, *Content and Consciousness*. S. Stich's *From Folk Psychology to Cognitive Science* pursues the matter further and expounds as well as criticizes alternative accounts of intentionality, including the view that there is a language of thought written into our brains. This view is defended by J. Fodor, especially in *The Language of Thought* and *A Theory of Content*. By contrast, John Searle, *Intentionality*, draws conclusions about belief and desire very different from those drawn by these philosophers. So does Thomas Nagel, in *The View from Nowhere*. Authors of both works argue against a naturalistic approach to reasons. An important recent defense of naturalism about reasons with a naturalistic account of intentionality is F. Dretske, *Explaining Behavior*.

Students interested in the debate in the philosophy of psychology about the character of folk psychology should consult Scott Christensen and Dale Turner, *Folk Psychology and the Philosophy of Mind*. A selection from this anthology of particular interest in the present context is "Folk Psychology Is Here to Stay," by Terence Horgan and James Woodward, which argues that the theory is compatible with a thoroughgoing causal cognitive science. S. Stich and T. Warfield, eds., *Mental Representation*, is an excellent collection of papers addressing the philosophical problem of how it is possible for the brain to represent the world.

CHAPTER FIVE

Behaviorism in the Behavioral Sciences

The most influential twentieth-century attempt to circumvent the difficulties that the mind/body problem raises for social science was behaviorism. This chapter traces the important motivations for this research program, which influenced all the social sciences. The reasons for its eventual eclipse as a research program reemphasize the problem of how beliefs and desires explain, and how central that problem is to the understanding of human action.

BEHAVIORISM IN THE BEHAVIORAL SCIENCES

Behaviorism was a widely adopted label for a variety of social scientists' responses to the problems of intentional explanation described in the last two chapters. In the second half of the twentieth century many scientists attributed the difficulty of developing a predictively powerful science of human action to the inability to observe and control mental states—beliefs, desires, etc. Inspired by the logical positivist demand that science base itself on what we can observe and control, many social scientists determined to restrict their explanations of behavior to those factors that can be observed: environmental factors that modulate, modify, elicit, trigger, or otherwise bring about the behavior to be explained. These social scientists, in psychology, sociology, political science, and elsewhere in the study of human affairs, called themselves behaviorists.

What behaviorists seemed to agree on was that the test of good social science should be a predictively successful, explanatory unification of observable behavior, whether of individuals or of groups. They held that social science should at least aim for consistent improvement in that direction.

Behaviorists held that taking intentional explanation seriously made it impossible to meet this challenge. Some gave explanations of this fact rather like (though less sophisticated than) the ones offered in Chapters 3 and 4. But most behaviorists simply expressed skepticism about theory in general. Others were explicitly dubious about "mentalistic" hypotheses, ones that involve attributing undetectable mental mechanisms to people, not just because there seems to be no way to test such claims directly and independently. Some behaviorists seemed to think explanations of action in terms of beliefs and desires were rationalizations or even confabulations irrelevant to the causes of the behavior. Other sophisticated behaviorists found intentional states especially objectionable because they could not see how mental states, especially conscious ones, could be physically embodied causes of behavior. All behaviorists held that it is such physical causes that we should seek.

For about twenty-five years after the middle of the twentieth century, the term *behaviorist*, as well as its variant, *behavioralist*, was a fashionable one. But like *positivism*, it came to have an unfavorable connotation after the 1970s. Though many social and behavioral scientists continue to endorse its philosophy of science, few any longer accept the label. We shall use it here for convenience, and without any pejorative suggestion. Although behaviorism went into eclipse in almost all parts of social science by the 1990s, the problems of explanation and prediction to which it was a response remained serious, and the reasons for its failure are highly instructive. In exploring solutions to the continuing problems facing a predictively improving social science, we need to be aware of why behaviorism failed. What is more, if behaviorism is a dead end and there are no alternatives to the sort of science its proponents sought, the prospects for a noninterpretative, predictively powerful social science may well be dim.

Not all behaviorists used that term to describe themselves, and some would not have accepted the label. But it is not hard to identify movements in each of the social sciences that endorsed the position described here as behaviorist.

In experimental psychology, behaviorism was a well-known label. As a methodology, behaviorism dates back to the early twentieth century, but its most visible exponent was B. F. Skinner. Following Skinner, psychological behaviorists held that the aim of their science was not to understand the mind but to systematize observable behavior. Systematizing behavior means providing models, general statements, and eventually laws that enable us to correlate observable environmental conditions with the behavior they trigger. Behaviorists like Skinner had specific arguments that this systematization cannot be accomplished by any intentional theory. Important social psychologists and behavioral sociologists adapted the theory that psychological behaviorists formulated to deal with general social phenomena.

In political science the label was slightly changed: behavior*al*ism was widely held to be a revolutionary development with much the same goals and effects in the study of politics as it had in psychology. Behavioralists advocated substituting the study of observed political behavior for the study of political institutions through the documents that codify them. In general they were more concerned with explaining what people do than with what they say they do. The behavioralists' aim was generalization with predictive power. Of course many behavioralist political scientists had no animus against intentional theories to account for the data their research generated. But that is because, as we shall see, behaviorism in political science—and economics—often employed an intentional theory without taking it seriously.

In economics and the part of political science that takes its methodological inspiration and its theory from economics, behaviorism has two forms. Sometimes it embodies a certain interpretation of a part of folk psychology—the theory of rational choice—that purports to circumvent the difficulties intentionality raises. Sometimes behaviorism in economics excludes from the intended domain of its theory the actions of individual economic agents altogether, so that the problem of explaining their actions is not one economics need deal with at all.

The origin of behaviorism as a methodological response to problems of explaining human action was, of course, in psychology. Therefore, let us begin by sketching its claims.

CAUSATION AND PURPOSE

Behaviorism began with despair about the limits of folk psychology and other intentional theories that attempt to improve upon it. Such theories all too often substitute jargon terms like *drive* for the ordinary notion of "desire" or "want." But it turns out the substitute terms mean exactly the same as the original ones and make no improvement in the explanatory powers of the theory. Often theories inspired by folk psychology rely on introspection—looking into our own minds—for hypotheses that explain our actions and for evidence testing such hypotheses. But, the behaviorist argues, introspective testing makes the "mind" the judge, jury, and executioner of its own theories about itself, when in fact we know almost nothing about our own minds and nothing directly about other people's minds.

Behaviorism takes seriously the philosophical problem of "other minds," the question of how we can know the private mental states of others, if all we have access to is their behavior. The philosophical skeptic challenges us to prove that others even have minds: "For all I know, everyone else might be a robot." Behaviorism resolves this issue by arguing that for purposes of psychology we don't need to solve the problem. We don't need the hypothesis that people have minds—mental states like belief and desire, sensations like feeling pain or noting colors. The reason we have no need of such hypotheses is that psychology is not the study of the mind. It is the science of behavior.

One argument for this view is known as *philosophical behaviorism* because it is a thesis about the definitions of psychological terms. According to this view, even statements that look as if they are about the mind are really disguised claims about behavior or are translatable into statements about behavior. Thus, when we say someone believes it will rain, that is not a report of some inaccessible inner state; rather, it means that he will carry an umbrella or a raincoat or stay indoors, et cetera.

But even if we allow an "et cetera" in a definition, this one will not do. For the belief will be expressed by such behavior only if the agent wants to stay dry. The definition must include, along with descriptions of behavior, a statement about the person's wants. But as we saw in Chapter 4, what a person wants turns on other desires and beliefs. So, the clause about a person's beliefs that is required to make the definition correct reintroduces the mental states the behavioral definition sought to remove. Thus it deprives the definition of its behavioral character.

The idea that we can translate away statements about the mind into statements about behavior did not last long among either philosophers or psychologists. It's worth mentioning only to distinguish it sharply from "psychological behaviorism," a far more substantial doctrine. Psychological behaviorism does not deny that there is a mind. Rather, it declares that questions about the mind are irrelevant to scientific psychology. That is because, first, human behavior can be explained without appeal to the mind; second, it cannot be explained by such appeals; and third, questions about the mind are themselves unanswerable in any case. So, for purposes of science, we might as well just ignore them and the mind.

The upshot is that intentional hypotheses like [L] cannot explain human actions because [L] is "mentalistic," that is, it refers to mental states of belief and desire. Behaviorists like Skinner are generally in agreement with the reasons discussed in Chapter 4 for [L]'s failure as a causal hypothesis. But they locate the problem at an even deeper level than the difficulty of testing [L] or improving its predictive power. Behaviorists view intentional explanations as the last refuge of an outdated form of scientific theory. This theory, according to behaviorists, started out with Aristotle in the fourth century BC as the ruling explanatory strategy for everything. But it has been shown by five hundred years of scientific advance to be incapable of really explaining anything.

In Aristotle's physics, his biology, and his explanation of human behavior, the crucial notion is that of purpose or goal. The behavior of all things is to be explained in terms of the purposes it serves. Thus, objects fall to earth because they seek their "natural place"; birds migrate to survive the winter; and people do the things that serve their ends. Such explanations are called teleological—from the Greek *telos*, meaning "end" or "goal."

We can see the difficulty of such explanations: they involve later, future states of a thing explaining its earlier, past states. Such an explanation cannot be causal. For the future cannot cause the past. Therefore, how future effects explain past causes is a mystery. One traditional solution to this mystery among "prescientific" peoples was to extend the desire/belief model of human action to the explanation of natural processes. Explaining human behavior in terms of prior states of belief and desire seems to avoid the problem of future goals explaining past acts. In the human case the goals are "represented" in the mental states that cause the acts.

So one way to underwrite teleological explanations of why rocks fall to earth or birds fly south is to appeal to God as an intelligent creator of the universe and to explain what happens in the world in terms of his purposes, his desires. Thus, birds fly south because God designed them that way, and he did it so that they would survive the winter. The purpose that flying south serves isn't the birds'—it's God's. It is relatively easy to explain anything that happens by this strategy: "It's God's will."

One problem with such explanations in terms of the purposes of the creator is to identify exactly which purpose is served by any particular natural process. For example, William Harvey's discovery that the function—the purpose—of the heart's beating is to circulate the blood required careful laboratory investigation and theoretical hypothesis formation. Teleological explanations thus leave lots of real work for the scientist. But the work has to be harnessed together with assumptions about God, or some intelligent designer of the universe, in order to have any explanatory power.

The history of natural science reflects the persistent shrinking of the domain of this explanatory strategy. It's not just that atheists, who do not believe in an intelligent designer, objected to explanations that required God in order to work. It was scientists who did believe in God who gave up this explanatory model. They discovered that they could provide more accurate and more powerful explanations of phenomena that were not teleological, that didn't involve attributing purposes to anyone. These discoveries started with Galileo, Kepler, and Newton, who showed that the motion of objects could be explained and predicted to extraordinary degrees of accuracy by appeal to their position, mass, acceleration, and gravity, without any reference to what purposes their motions might have served. This kind of *mechanistic* theory has no need for teleology. It has been the dominant explanatory strategy in physical science to the present day.

Mechanism has been preferred to teleological explanations, not because it needs no hypothesis about God or about inanimate objects having purposes. Rather, it has become dominant because mechanism is so much better at predicting phenomena than teleological theory. After all, even if we can be sure what God's purposes are, we still can't tell how he will attain them—exactly what behavior aimed at this purpose will be produced. For example, if God wants birds to survive the winter, there are lots of ways he could have arranged it besides having them fly south in September—a thick fur coat, hibernation, or a milder winter in the north would all do the trick. Once we see what animals do, we can explain it after the fact. But before we see what they do, knowing God's purposes doesn't narrow down the alternatives enough to enable us to make a prediction.

But, of course, since Newton, few scientists of any kind have been willing to appeal to God to explain phenomena, which has long made teleological explanations even more problematic. That is especially true in biology, where teleological description and explanation are impossible to eliminate. Biology begins with the identification of functions: wings are for flight, the liver stores sugar, and so on. But functions are just the things that organs do *in order that* the animal can move or breathe or store energy, for example. That is, functions are things organs do to serve purposes. But whose? In the absence of someone's intentional states, there seems no basis for identifying the purposes and functions served by, say, an organ's behavior. And yet it seems undeniable that biological behavior serves purposes. Once God's intentions are ruled out, biology needs an alternative explanatory strategy that makes no appeal to teleology or a theory that provides a purely causal underlying mechanism for biological teleology. Charles Darwin's theory of natural selection is sometimes said to have been the former and sometimes the latter. In either case it solved the biological problem of teleology.

Darwin's theory teaches that the purposive appearance of biological phenomena is the result of a large amount of blind, unpurposeful, heritable variation. In an entirely random way, many different heritable traits, including behavior, are produced in every generation, and nature—the environment—selects the fittest of them. By fittest, Darwin meant the likeliest to survive and enhance the organism's chances of reproduction. Thus, flying south came about because the birds that just happened to be genetically programmed to move away from cold weather survived and reproduced. Those not so programmed didn't fly south and didn't survive. And how did the genetic programming for flying south when it got cold arise? It arose, along with many other less adaptive behavioral traits, through the random recombination of genes, or mutation, or some other entirely causal, nonteleological process.

Some biologists and philosophers argue that Darwin's theory showed that teleology is just as unneeded, indeed, just as unscientific, in biology as it is in physics. With this view, when we talk about biological functions, we are not really describing purposes or goals. We are just invoking shorthand for Darwinian adaptation—variation and natural selection. There are no real purposes in nature. Biology is thus every bit as causal as physics. Other biologists hold a less radical thesis. They claim Darwin's theory didn't explain away the appearance of teleology; rather, it justified teleological explanation in biology. For the natural selection of blind variations underlying such explanations is a purely causal mechanism. Thus, Darwin is held to have naturalized purpose—reconciled its existence with the purely mechanical processes we recognize in physics.

We don't need to settle the dispute about whether Darwin eliminated purpose or naturalized it. For our purposes, the moral of the story is that the appeal to intentions has been as ruthlessly read out of biology as it was cast out of physics. Both the life sciences and the physical ones are thoroughly causal disciplines, committed to the search for laws that will provide causal explanation.

That leaves only the social sciences as the last refuge of an explanatory strategy that started, over 2,000 years ago, as the ruling paradigm in all science. The behaviorist wants to rid social science of purpose as well, to substitute causal theory for teleological theory on its home territory. Why? It is because of a conviction that appeal to intentional states won't provide for human behavior the kind of causal theory the rest of science aims at.

In fact, despite the appearance of being causal, the explanations of folk psychology are really teleological after all. Desires and beliefs may be causes, but it is their content—the statements they are about—that does the explaining in an explanation of action. And these statements describe future purposes, not prior causes. First of all, behaviorists argued, folk psychological explanations cannot be underwritten by appeal to an underlying causal mechanism like the one Darwin provided for biological explanations. Recall the problem of Chapter 4: we cannot identify beliefs and desires in terms of brain states or other signs independent of their effects, which are actions. Desires and beliefs look like distinct prior causes of action, but there are no descriptions of them distinct enough from their future effects. Therefore, behaviorists argued, such explanations cannot be informative: they do not describe causes in terms logically independent of our descriptions of effects.

The behaviorist claims that intentional explanations are uninformative because they are disguised teleological explanations. To see that, suppose I explain someone's going to a food store by pointing to his desire to get food and his belief that food is available in the store he's heading for. The only way I can identify his desire for food is by citing the food he eventually gets. Therefore, the cause of his behavior, the desire to get food, is identified only by reference to an event that occurs later than the action his desire explains: heading to the store. This reveals that the apparently causal explanation, in terms of prior desire, must make covert reference to a future event, getting the food that satisfies his desire, in order to explain an earlier event, his going to the store. The apparently causal explanation is really a disguised teleological one. And the teleology cannot be eliminated because the only way to identify his desire is in terms of the event that later satisfies it. Of course, sometimes desires are unfulfilled. For instance, our hungry person may be run over on the way to the store and never fulfill his desire. But that makes matters worse. For now if his going to the store is explained by the desire to get food, it is ultimately explained by an event that never happens at all, his getting the food.

Therefore, intentional explanations are at least covertly teleological, behaviorists argue. And in their objections to teleology, history is on their side. The replacement of teleological forms of explanation by causal ones has been a consistent trend over the past four hundred years. Not only that, but it has been a trend with big payoffs for the increase of explanatory depth, unification, and predictive precision that teleology cannot obtain. So, it should make sense to effect this replacement of causes for purposes in the social sciences as well.

Some philosophers and social scientists will add philosophical arguments to the historical one. They read the historical record as vindicating a metaphysical worldview known as *physicalism*. That is the idea that the world is composed of nothing but matter in motion, from quarks to stars, matter behaving in accordance with the laws of physics, which, together with initial conditions, determine the world's future completely—or if quantum mechanics is right, to precisely determined levels of probability. This is a world with no room for teleology and with no room for irreducible intentional states. Our belief that there are such states is an illusion, fostered by a misleading reliance on introspection. Behaviorists do not openly acknowledge such a view, for they claim to be agnostic about matters of metaphysics, which have no direct relevance for observation. But to the extent that a scientific method and the arguments for it do reflect metaphysical commitments, physicalism is a view behaviorists should be comfortable with. Psychology is thus the science of behavior, because in the end, that's all there is for any science to study: matter in motion—behavior.

Behaviorists rejected teleology and sought a research strategy that would avoid it. The one they hit on did not work, as we shall see. In fact, it can be accused of smuggling in the very teleology they rightly rejected. Behaviorists, like other empiricist scientists and philosophers, were right to reject teleology, and right for the reasons that they gave. Wherever the appearance of purpose unexplainable by nonteleological, causal processes arises—in the social sciences or in commonsense explanation—we need an explanation of how it is possible. For we know that the future cannot cause the past and that invoking God's design is never an acceptable explanation. This problem is particularly serious in the social sciences, for in almost all of them descriptions and explanations of the emergence and the behavior of social institutions accord them *functions*, that is, purposes, usually ones that are not the same as the purposes of the human beings who participate in these institutions. The problem of underwriting and justifying such functional, purposive explanations will be the focus of Chapters 9, 10, and 11.

THE EXPERIMENTAL ANALYSIS OF BEHAVIOR

What theory of human behavior do behaviorists offer in place of folk psychology and its reliance on [L]? One alternative ruled out by most behaviorists is a theory that explains human behavior in terms of neurological events and processes in the brain. The reasons for this elimination are not philosophical; they are purely tactical. We could, to be sure, eventually provide the neurological causes for any particular action of any particular person at any particular place and time. But for the purposes of social science, this explanation would be useless, even if we were prepared to wait for the time in the distant future when neuroscience will be able to provide this information.

The fine structure of the brain differs so much among people that the exact details of our neurological explanation in one case would probably not be applicable to anyone else doing exactly the same thing—saying, "Please hold the door," for instance. But these details are just what are crucial to neurology's claims to improve on the predictive weakness of folk psychology. As scientists, psychologists want to explain kinds of behavior, not individual instances of it. Explaining individual instances is crucial to testing our theories. But predicting future events requires laws and theories about the "natural kinds" these future events will exemplify. The kind-vocabulary of neuroscience will include synapse firings and acetylcholine production, but it won't include "deciding to vote Democratic," "preferring coffee ice cream over vanilla," or "believing that eleven is a prime number."

The behaviorist aims to develop a theory of behavior that can be systematically linked up with neuroscience. In fact, in the ideal case, the theory's laws and theories will themselves be explainable by neuroscientific ones, in the way that the behavior of chemical substances is explained by the laws governing the atoms that compose them. But this "reduction" of psychology to neuroscience is a long way off. It will never be accomplished anyway, unless behavioral psychology first finds some laws and theories that will supersede and improve on folk psychology. Otherwise, there is nothing to link to neuroscience.

Behaviorists hold that there is another reason we can neglect neuroscience in the short run. They say that behavior, both human and animal, is in fact largely a function of environmental factors alone; at any rate, we can ignore the intervening neural details and still explain almost all of what interests us in human behavior. For there are predictively powerful explanatory generalizations about behavior that link it directly with observable environmental variables. In fact, behaviorists like Skinner and his followers lay claim to having formulated a generalization of this sort to replace our intentional law [L]. This is the "law of effect," the leading principle of "operant behaviorism." Examining the law of effect will be a convenient way to assess the claims and prospects for behaviorism as a more scientific alternative to folk psychology.

Let's return to the distinction that specifies the domain of social science, the difference between "mere" behavior and action—the product of beliefs and desires. According to behaviorists, the distinction between mere behavior and action needs to be replaced, just as Aristotle's distinction between rest and motion was superseded by Newton's distinction between rest (including rectilinear motion) and acceleration. Skinner's distinction is between "mere" behavior—roughly, reflexes like blinking—and what he calls "operant behavior." Reflex behavior can generate relatively complicated results, provided that it is properly linked to environmental conditions. Recall the famous experiments of the Russian psychologist Pavlov: by associating the sound of a bell with the presentation of food, Pavlov was able to get dogs to salivate when the bell rang alone, in the absence of food. However, it's clear that little of what we call human action can be the result of conditioned reflex.

But, claim behaviorists, complex human actions can be the result of conditioning operant behavior. Operant behavior is behavior emitted by the organism as a result of neural causes that are complex, varied, and too irregular to be of interest at the present level of neuroscience. Thus, operant behavior differs from reflex behavior, which is produced at least originally as the routine result of a single, easy-to-identify neurological reaction to a single, easy-to-identify stimulus. Operant behavior is the sort of thing we might describe in ordinary terms as “voluntary.” But once emitted, the more complicated bit of behavior becomes subject to environmental conditioning, in accordance with the *law of effect*:

> [LE] If emitted behavior is reinforced, it will be repeated with greater frequency (or intensity or duration). If it is punished, it will be repeated with lower frequency (or intensity or duration).

Roughly, the law of effect claims that once emitted, for whatever cause, a bit of behavior is likely to be repeated (or otherwise strengthened) if it is followed by some sort of benefit—the reinforcement—to the organism. Behavior whose frequency, intensity, or duration can be increased by such reward or reinforcement is operant behavior. The term also covers behavior that, once emitted, is reduced in frequency if it is followed by some sort of loss or cost to the organism. All human actions, on this view, are instances of operant behavior. The rate of occurrence of any particular action can be explained by the frequency with which similar actions have been reinforced or punished in the past.

Skinner first exploited the law of effect in the explanation of the relatively complex behavior of rats and pigeons. By providing them with food pellets after the emission of some behavior like bar pressing or key pecking, he was able to get them to emit this behavior in a predictable fashion. By varying the schedules of reinforcement, he was able to vary the frequency of the behavior and to control other aspects of it. By reinforcing successive refinements of emitted behavior—*shaping* through *response differentiation*—he was able to get the animals to undertake complex and highly unnatural movements and to perform remarkable feats of discrimination. A pigeon can be trained in this way to discriminate among dozens of different geometrical shapes or to dance around its cage. The association of an environmental feature with a reinforcing stimulus—*secondary stimulation*—enables factors only distantly connected to the direct reinforcement of the organism to control operant behavior. For example, monkeys can be trained through the right sort of reinforcement to respond to tokens or chips—money—that they can exchange for direct reinforcers like food.

Skinner was able to replicate his experiments in a way that convincingly confirmed the theory with regard to the behavior of these animals. What is more, by identifying and employing appropriate reinforcers, by shaping and secondary stimulation, psychologists have wrought remarkable changes in the behavior of people. For example, psychiatric patients previously thought uncontrollable can be made to look after many of their basic needs by “token economies” of the sort successfully employed with monkeys.

Skinner was certainly not timid in his claims about all human behavior: everything we identify as action is scientifically to be explained as operant behavior. Everything we do is under the control of the complex and varied arrangement of reinforcements in our environments. Subtle and sophisticated behavior, like speech, for example, is the result of shaping and secondary reinforcement. Of course, when adults teach babies to talk, they do not realize that they are engaging in operant conditioning. Parents do not view their behavior as setting up schedules of reinforcement, shaping, and associating secondary reinforcers in such a way that a child successively refines emitted noises into speech. Once a child has learned to speak, its verbal behavior continues right through adulthood to be subject to reinforcement and punishment. Like the rest of human behavior, speech is under environmental control, where the environment includes both nature and other people.

One thing stands out about this theory. Despite Skinner's claims, it is a thoroughly teleological one. It explains a bit of behavior in terms of apparently subsequent reinforcement. For example, people speak grammatically because they are reinforced after doing so (by smiling responses) and punished after failing to do so (by grimaces and corrections). Here we have a claim just like the biological assertion that the heart beats because its beating is followed by the blood's circulating.

But, the behaviorist claims, this is entirely innocent teleology, teleology naturalized by Darwin. Or perhaps it is only apparent teleology. For it is either shorthand for or is backed up by an entirely causal theory. Indeed, exactly the same type of theory that backs up the functional analysis of the heartbeat underwrites operant conditioning. It is another form of natural selection. Just as successive hereditary variations are shaped by selection for their consequences, so successive bits of behavior are selected for their consequences—reinforcements. In evolution, the consequences of selection are increases in the population of the species. In individual behavior, the consequences are the increases in the repetition of the behavior. "My heart beats in order to circulate the blood" means roughly that my heart beats because hearts have been selected in the past for their efficiency in circulating the blood. "I speak grammatically because this behavior is reinforced" means that speaking grammatically has been selected for (reinforced) in the past, and therefore it recurs now.

In both evolutionary and operant selection there is a problem about underlying mechanisms. How does the selection of hearts for circulation in the distant past get connected to my heart's present behavior? Natural selection is silent on this question. But the answer is provided by genetics and physiology. The genes bear the program of the heart's structure. Nature didn't select the most efficient hearts directly; rather, it selected the genes that encoded successively more efficient heart structures. These genes are passed down by reproduction, thus providing the mechanism connecting the evolutionary past to the present.

Operant behaviorism requires a similar mechanism. Presumably, this mechanism is ultimately to be given by neuroscience. Somehow the reinforcement of earlier behavior rearranges the neurology of the individual in such a way as to increase the frequency of the behavior's occurrence later; similarly, punishment of behavior—the opposite of reinforcement—must be based on a neurological process that extinguishes the behavior. Indeed, a start has already been made on uncovering these processes in the neurology of the sea slug. But evolutionary biology can do some things independently of the genetic basis of underlying mechanisms. Similarly, it is argued, experimental psychology can do much without the mechanism that neuroscience must ultimately provide for it.

Operant behaviorism has been an outstanding success in the laboratory: though it has serious limitations, its powers to predict and control behavior in the stereotyped conditions of the experimental analysis of animal behavior are undeniable. However, its application to human behavior, to the domain of folk psychology, has been far less successful. Indeed, it seems no more powerful in its ability to explain and predict human behavior than intentional theories have been able to predict human action. That may simply be because operant psychology is a young science. Perhaps more research is needed to parlay its undoubted successes with laboratory animals into an equally successful behavioral approach to people. As a young science, it needs to be given a chance, claim behaviorists. At any rate, it does not have the obvious methodological defects of folk psychology and its intentional successors.

THE GHOST IN BEHAVIORISM'S MACHINE

In fact, whatever plausibility behaviorism has is said by some to derive from its being old wine in new bottles. For behaviorism may turn out to be nothing more than folk psychology translated into new jargon. Critics of behaviorism argue, first, that theories of behavior are just as teleological as the intentional theories they are meant to supplant and, second, that these theories turn out to be intentional ones themselves, so behaviorists are really just fooling themselves about their "alternative" to folk psychology.

The basis of the charge is the suggestion that when organisms, rats or people, emit certain sorts of behavior, it is because they *want* to be reinforced for it and *believe* that they will be. Behaviorists resist this suggestion. But it is easy to see how an operant explanation of a bit of behavior can dovetail with an intentional one. This dovetailing of [L] and [LE] is a strength and a weakness. It is a strength because it makes clear that any behavior explained by [L] can be explained by [LE]. Thus, the law of effect bids fair to replace folk psychology's outmoded model without any loss.

[L] and [LE] dovetail so nicely that at various times it has been tempting to explain one in terms of the other. Advocates of intentional psychology will explain why reinforced behavior is emitted by pointing out that people (and maybe animals too) like reinforcements and will do what they believe necessary to secure them. Behaviorists can run the explanation in the opposite direction: people want certain things and do what they believe will secure them because those things are reinforcing. But behaviorists are not inclined to offer this explanation of folk psychology, because they consider [L] to be without explanatory merit. Furthermore, they consider the intentional concepts in which it is expressed to be unscientific. Behaviorists don't want their explanatory notions to be closely related to those of folk psychology, because that would tar them with the brush of its failures.

The dovetailing of [L] and [LE] is a weakness because it reveals that [LE] may face the same problems as [L]. For [LE] to be superior to [L], we need to identify *descriptions* of reinforcers and stimuli independent of the behavior they control. Otherwise, the reinforcers and stimuli will be in the same boat as wants and beliefs. Unless behaviorists can make such descriptions, it will be very difficult to show that operant theory is really different from folk psychology or that [LE] escapes the intentional character of [L]. If it doesn't, then the excuse for behaviorism's present predictive weakness—that unlike folk psychology, behaviorism really is a young science—will turn out to be a hollow one.

In fact [LE] turns out to have all the methodological infirmities of [L]. And the simplest explanation for this fact is that the law of effect, [LE], is just [L] dressed up in new and more scientific-sounding terms.

To distinguish itself from folk psychology, operant behaviorism must do two related things: First, it must provide a means for identifying environmental stimuli, reinforcers, and operant behavior independent of one another. Second, it must demonstrate that none of these three factors are directly or indirectly intentional. That is, it must show that they don't involve "content" or representations of the way the world is or could be.

Let's start with behavior. Behavior under the control of operants is "learned" behavior. Behaviorists define learning as simply a change in the response rate to a certain stimulus. But what is learned? Take the case of a rat in a maze. When a rat learns the maze's configuration so that it takes a direct route to the food, what is the behavior it has learned? Is it a series of steps, twelve steps of two centimeters each straight ahead, four centimeters to the right, two to the left, and so on? No, for the rat does not always take the same number of steps of the same stride to get to the food. Sometimes it stops to scratch, sometimes it runs, other times it walks in a leisurely way. The rat never takes exactly the same route at the same speed twice. Yet all these behaviors are examples of the same learned behavior as long as it gets there fast enough. What do they all have in common? What makes them all instances of the same learned behavior? It cannot be that they are the same because they all result in eating, for sometimes the experimenter removes the food and yet the same behavior is produced. It seems unavoidable to say that the behavior the rat has learned is food-finding behavior.

The behavior the rat learns is thus defined by reference to its goal. But first, this definition is improperly teleological; second, the food is sometimes not there to be found, yet the behavior is the same. The obvious solution to this puzzle is to say that what the rat has learned isn't a set of movements with some end, goal, or purpose. What it has learned is where the food usually is. But that means it has learned something about the world—something that can be expressed in a proposition ("the food is at the end of the left alley"). To have been learned by the rat, this proposition must be *represented* in some state or other of the rat. We may not want to call this representation in the rat's brain a state of belief about where the food is. But it is still an intentional state of some kind. If that is the only way to identify the behavior that the rat has learned, then we are in exactly the same boat with respect to operant behavior as we are with respect to action.

The same suspicion arises for the notion of an environmental stimulus, an observable feature of the environment leading to a operant response that can be reinforced. In the case of a pigeon, the stimulus may be a button: the pigeon is reinforced for pecking it when it is lit, but not otherwise. In the case of animals under experimental conditions, it is relatively easy to identify stimuli. They are changes in the environment that are correlated with reinforced responses in accordance with the law of effect. If we can train a pigeon to discriminate shapes, then shapes are environmental stimuli. Suppose we try to train a pigeon to discriminate odors, say, by reinforcing key pecking whenever ammonia is presented and never reinforcing key pecking in the presence of gasoline. If we fail to get it to peck in the presence of ammonia or to avoid pecking in the presence of gasoline, then these odors are not stimuli for the pigeon. But notice, we are now using [LE] to test [LE]. That is, we set out to tell whether the pigeon can discriminate ammonia from gasoline. The way we do that is by applying the law of effect to ammonia and gasoline in order to see whether it controls behavior when paired with reinforcers. If one of them does, then it's a stimulus. If it doesn't control behavior, then it isn't a stimulus. [LE] works with stimuli and reinforcement to cause behavior. And to determine what range of environmental factors can be discriminated by an animal, we appeal to [LE].

As we've seen with [L], it is permissible to use [LE] to test itself (though it weakens the testability of [LE]), provided we have some alternative means of identifying environmental stimuli. This proviso will be especially important when we leave the laboratory and turn our attention to human behavior. What are the alternative means, besides [LE], of identifying stimuli? Obviously something is a stimulus if the organism can perceive it, see it, hear it, taste it, feel it, and so forth. But that can't mean simply that photons from the object falling on the retina or waves of air pressure that strike the inner ear are automatically stimuli. Lots of things impinge on an organism's sense organs this way without being registered. To see a red apple is not simply to have the required peripheral sensations. Folk psychology tells us that seeing the apple involves having the concepts *red* and *apple* and classifying the sensations under those concepts. But that looks like another case of bringing something under a description, that is, coming to have something like a belief about it. We are again faced with a choice between intentionality or circularity. For suppose behaviorists argue that seeing a red apple—having the stimulus—is just overtly responding in an appropriate way (say, reaching for it and taking a bite out of it). Then they have not provided an alternative means of identifying stimuli; they have merely redescribed it in terms of what the operant behavior [LE] tells us the stimulus "controls." And if they admit that seeing a red apple is bringing it under a description, they embrace the very sort of intentional theory behaviorism repudiates.

When we try to understand reinforcement, the role of covert intentionality is perhaps clearer. A reinforcer is whatever changes the organism's response rate. Again, in the laboratory, it is easy to vary the presence and absence of conditions to learn whether they are reinforcing. The results are quite remarkable, for the range of reinforcers is very broad. Food is a reinforcer, of course (when the organism has been deprived long enough). But saccharin, too, or even the opportunity to look out of a window is a powerful reinforcer for monkeys. Indeed, monkeys have been trained to solve puzzles when the only reinforcement present seemed to be solving the puzzle. And the range of items humans find reinforcing must be staggering.

What do all these reinforcers have in common? What makes them all reinforcers? The tempting answer is that organisms—animals and people—want them, desire them, find them satisfying, and so on. But that is not a permissible answer for the behaviorist, for it rests on an intentional notion of want or desire. Of course, somewhere deep in the brain, all reinforcers must have some small number of neural features in common. In fact, behavioral psychologists were eventually able to combine [LE] with rudimentary neuroscience. It was noticed that rats implanted with electrodes at the midbrain would literally work themselves to exhaustion if a bar they could press closed an electric circuit that stimulated the part of their brains with the electrodes. James Olds, the psychologist who undertook this particular experiment, hypothesized that all reinforcers operate through connection with this portion of the brain, which he called the pleasure center. But this hypothesis could not immediately provide a useful means of identifying what in general all reinforcers have in common besides apparently doing things to our brains—something presumably we knew already. Having a neural effect in common is not a feature reinforcers have that would enable us to identify them independently of the behavior they bring about. At least it won't do so until neuroscience provides some litmus test for this shared neural effect. Therefore, neuroscience was not able to help operant behaviorism improve on folk psychology in the explanation and prediction of human behavior. Even now, long after the eclipse of narrowly behaviorist research programs, and in a period of rapid expansion of cognitive neuroscience, psychologists remain unable to improve very much on folk psychology in the explanation of normal behavior. At this point neuroscience has begun to outline the physiological capacities and incapacities that actual behavior reflects. But it cannot yet explain exactly what we do from action to action and why we do it. Doubtless neuroscience will continue to enhance our diagnostic powers, especially in regard to psychopathology, but its bearing on the sort of action and behavior that concerns the social sciences will be at best limited for a long time. The brain is far too complex for neurological litmus tests that might allow for easy identification of the particular beliefs and desires that bring about actions.

CONCLUSIONS

Without denying the important accomplishment of behaviorism in the explanation of aspects of the behavior of laboratory animals, our conclusions about this theory as an alternative to folk psychology cannot be optimistic. The motivations of behaviorism in the history of scientific progress may well have been laudable. Its aim of avoiding intentionality may be the right one for providing a science of human behavior. But at a minimum, such a science appears more difficult to attain than the behaviorist supposes. For the obstacles the behaviorist must surmount to attain a predictive theory seem identical to the ones folk psychology apparently cannot overcome. And behaviorism seems too much like folk psychology to succeed where the latter has failed. That leaves the naturalistic approach to human action no better off than it was before the behaviorist offered to solve its problems. In Chapter 7 and especially in Chapter 8 we shall see how the antinaturalist deals with these problems.

Meanwhile, we need to trace out the influence of behaviorism in economics and political science. These are disciplines in which "behaviorism" is the label for a slightly different approach to the search for improvable laws about human behavior. We shall see that in these disciplines, "behaviorism" raises more issues for the philosophy of science than it settles.

Introduction to the Literature

The most vigorous exposition and defense of behaviorism is to be found in B. F. Skinner's *Science and Human Behavior.* A useful introduction to the experimental psychology influenced by Skinner is H. Rachlin, *Introduction to Modern Behaviorism*. Behaviorism in sociology is associated with the work of George Homans, *The Human Group*. See also an anthology of writings in the area, edited by R. Burgess and D. Bushell, *Behavioral Sociology*. J. C. Charlesworth, ed., *The Limits of Behavioralism in Political Science*, treats the impact of this movement. His anthology contains an influential paper by D. Easton, "The Current Meaning of Behavioralism in Political Science."

The two most sustained attacks on behaviorism are found in Charles Taylor's *Explanation of Behavior* and in Noam Chomsky's well-known "Review of B. F. Skinner, *Verbal Behavior*." The effect of these criticisms has been to encourage the transformation of behavioral psychology into "cognitive science." See D. C. Dennett, *Brainstorms*, and P. Churchland, *Matter and Consciousness*, for an exposition of these developments.

Important discussions of the nature of teleological analysis and explanation by philosophers are to be found in E. Nagel, *The Structure of Science*, Chapter 12, and C. Hempel, "The Logic of Functional Analysis," in his *Aspects of Scientific Explanation.* A criticism of their views is R. Cummins, "Functional Analysis." L. Wright, *Teleological Explanations*, is the most relevant analysis of teleology for the behavioral and social sciences. It offers an important counterweight to Taylor's argument in his *Explanation of Behavior*, that biological and behavioral teleology is both ineliminable and causally inexplicable.

CHAPTER SIX

Problems of Rational Choice Theory

Since the nineteenth century, economists had supposed that they were well on the way to turning belief/desire explanations of human action into a quantitative scientific theory: rational choice theory. The difficulties they faced turned out to be the same as those that common sense faces, and for at least some of them, the reaction was behaviorism. Despite its persistent and indeed increasing attractiveness among social scientists, the economist's approach to explaining human choice, or avoiding the need to explain it, faces all the problems of behaviorism, and some more of its own.

THE THEORY OF RATIONAL CHOICE

Economic theory and parts of political science employ a formal theory of rational choice. But, as we shall see, in certain respects rational choice theory is folk psychology formalized. Thus, these disciplines need to come to grips with the problems outlined in Chapter 4. Economics and political science face the same problems that bedevil other disciplines that employ [L]. It is to this employment of [L] that we may trace the recurrent criticism of economic theory as being either false or vacuous and untestable. The more accurate charge such criticisms reflect is that theories relying on [L] have some predictive power but are not improvable in explanatory detail and predictive precision. Unlike behaviorists in psychology, economists have not self-consciously sought a substitute for folk psychology. Instead, they have tried to make it more precise, and they have extracted from it some surprising and powerful consequences. Once behaviorism became fashionable in the behavioral sciences, economists tried to give rational choice theory a new behaviorist interpretation that would circumvent criticisms of its explanatory and predictive weakness. Economists' chief reasons for preserving rational choice theory have to do with the remarkable consequences they have drawn from it about large-scale social processes and institutions, consequences that make it seem indispensible, even if it is also empirically ungrounded. This combination of features makes rational choice theory a matter of enduring interest to philosophers.

RATIONAL CHOICE AND THE INVISIBLE HAND

The theory of rational choice is most easily understood if we approach it historically. In the nineteenth century, economists known as *marginalists* (how they got that name is explained below) systematized economic activity based on a theory of utility. They held that every economic agent could derive a certain amount of utility or satisfaction from any amount of any commodity. Furthermore, subject to the limitations on available resources and information, each agent acquired the bundle of commodities that maximized the agent's utility. Rationality was thus defined as the maximization of available utility, and all agents were assumed to be rational.

In effect, utility theory is a way of ordering desires. It enables us to dispense with the ceteris paribus clauses in [L] that implicitly exclude overriding wants. It is also supposed to provide a more precise quantitative theory of the strength of different wants and how they are combined and reconciled.

The marginalist economists and their successors typically assumed that agents have full information about the world, that is, they have true beliefs about all the facts relevant to their circumstances. With this proviso, the assumption of maximizing utilities might enable us to predict agents' choices from their utilities alone. Recall that given [L], we can derive action from desire if we can hold relevant beliefs constant. The full-information assumption is a means of holding these beliefs constant. For example, if a person knows what ice cream flavors are available and we know which one that person gets the most utility from, we can unerringly predict which one she will choose.

So the marginalists translated [L] into a theory about people's choices. As noted above, they held that people choose rationally: they always choose the available alternative that they believe maximizes their welfare, satisfaction, happiness, or pleasure, as measured in units of utility. From this theory, or hypothesis, model, axiom, or assumption, from this starting point, they were able to mathematically derive many other claims about the nature of an economy.

Economists originally limited the application of utility-maximizing theory to economic actions: market choices made in the light of costs and benefits, as measured by prices. But the mathematical clarity of this theory of choice eventually led economists and others to expand its range of application to many other actions explained originally by [L]. Social scientists have extended the notions of cost and benefit to ends and goals that don't have money prices but still involve trade-offs and therefore *shadow prices* and *opportunity costs*—the cost of one action as measured by the value of another action you must forgo to do the first one. Thus, if you have to choose between two potential marriage partners, the "cost" of one is determined by what you must give up in not choosing the other. Of course, such extensions of utility theory beyond the actual market to the implicit market in which all of our choices have prices (implicit or shadow prices) attached should be no surprise. For the range of rational choice theory is the same as that of [L].

The marginalists held that the utilities associated with commodities are measurable in real units that enable us to say how much more an agent prefers one alternative to another—twice as much, 20 percent more, et cetera. This view is called *cardinal utility* theory. The name derives from the hypothesis that utility, like mass (0 grams, 1 gram, 2 grams, for example), comes in amounts that can be measured in cardinal numbers: 0 units of utility, 1 unit of utility, 2 units. From the assumption of cardinally measurable utility, economists were able to derive important results about how the demand of individuals for commodities and the supply of commodities by producers varied with their prices. Suppose that the amount of utility derived from one additional unit of a commodity—the so-called marginal utility—declines as more units of that commodity are acquired. For example, if you crave an apple, the amount of utility derived from an apple may be large. If you have already had two or three, the amount of utility derived from the fourth will be much smaller. If marginal utility declines while the price of each apple remains the same, then the individual's demand for additional apples will decline. If you're hungry for apples, you will pay a fairly high price for the first apple. But if the price of the second apple is the same or higher than that of the first, you will be less inclined to buy it, for the utility it provides is less than the first one provides. The marginalists elevated this psychological proclivity into a law of declining marginal utility, hence the name *marginalists*.

The marginalists also showed that the amount of a commodity an individual will be willing to produce and sell—the supply—rises as its price increases, provided that the individual seeks to maximize profit. But that, of course, is just what a rational utility maximizer will seek to do if income from selling can be used to purchase goods that increase utility.

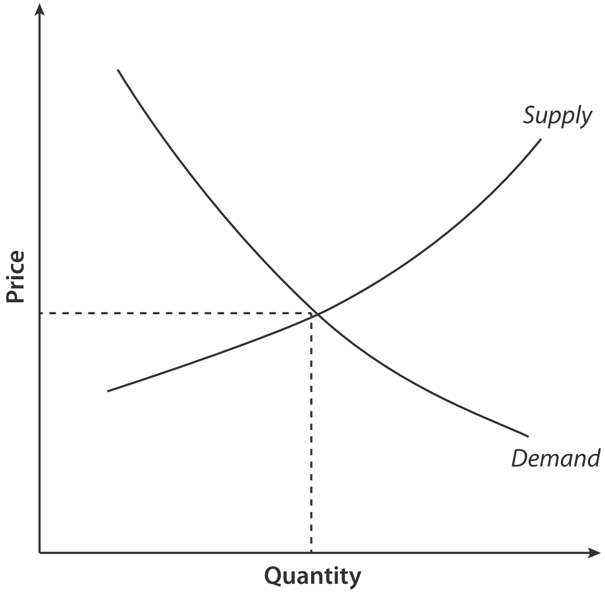

FIGURE 6.1 **Supply and Demand Curves Intersect at Equilibrium Point**

Now, for each individual supplier, we can plot the relation between price and supply on a graph, and for each individual demander we can plot the relation between price and demand on graphs (with price as the *x* axis and amount of each commodity demanded or supplied at each price on the *y* axis). These two curves will cross each other, the demand-price curve going downward and the supply-price curve going upward.

But the entire market for a particular good is just the sum of the supply and demand for that good by all the individuals in the market. By adding the individual supply curves and demand curves, economists can derive supply and demand curves for an entire market and, ultimately, the economy.

The supply and demand curves are, of course, familiar staples of economic theory. Their obviousness belies their importance, not only in economics but in many other theories that take rational choice seriously. The equilibrium resulting from the matching of supply and demand at the amounts and prices where the two curves cross reflects an explanatory strategy fundamental in many theories in social science—the so-called invisible or hidden-hand model that goes back to Adam Smith's *Wealth of Nations*.

Where these two curves intersect, there is what economists call a *market-clearing equilibrium*: market-clearing because when quantities supplied and demanded are in exact equality, there are neither shortages nor surpluses, no wasted production and no unmet willingness to buy more at the given price. At any other price, there will be unmet willingness to buy—that is, a shortage—or an unmet willingness to sell, an oversupply of goods that remain unsold—a surplus. Surpluses and shortages in markets are signals to buyers and sellers to change their plans. A surplus is a signal to producers to lower prices and cut down on production. A shortage is a signal to increase production and raise prices; similarly, a shortage signals to buyers that they need to increase the price they are willing to pay or reduce the amount they are prepared to purchase. Thus when the market price is different from the equilibrium price, the market sends signals to all participants about how they need to change their strategies of purchase and sale to maximize their utilities in the next round of exchange. Responses to these signals by utility maximizers result in the market's moving back to equilibrium, where there are neither shortages nor surpluses.

You might think the absence of any planning at all by government, or any agreement among all producers and consumers, would produce chaos, as well as continual shortages and surpluses. But as Adam Smith argued in the eighteenth century—about a hundred years before the marginal utility theorists—this supposition is in fact quite mistaken. Smith argued that a competitive market of rational individuals, all simply seeking their own self-interest, that is, maximizing their own utilities, coordinates every rational buyer's and seller's plan effectively and continually. The hidden or invisible hand of the market, by sending signals to rational agents, continually adjusts their actions to bring about a result that minimizes shortages and surpluses, that reduces the waste of resources and benefits everyone, *even though no individual buyer or seller intends, designs, or even expects this outcome.* That is why Smith described the operation of the market as the activities of a hidden or invisible hand. It produces an outcome that no one intended but that makes everyone better off.

Inspired by Smith, the marginalist economists and their successors set about attempting to turn his verbal argument into a mathematical proof of a scientific fact. Their successors eventually succeeded in proving something like this result. What they proved was that under certain assumptions, exchange between rational agents is guaranteed to produce a unique, stable market-clearing equilibrium. Moreover, it is a single global equilibrium, not several local equilibria. It is also one that is stable in the sense that when for some extraneous cause prices move away from it, the market inevitably forces prices back to the original equilibrium. And it has the no-waste, welfare-optimizing feature that there is neither a surplus nor a shortage at the unique, prevailing stable price. What is more, they proved that such a market will make people as well off as they can be made, all things considered, thus vindicating Smith's invisible hand. These mathematical results are known among economists as the general equilibrium theorems of welfare economics. It took more than 150 years for mathematical economists to finally prove this result and thus vindicate Smith's verbal but not mathematical argument. The mathematical economists who accomplished this feat were awarded the very first Nobel Prize in economics.

The successful mathematical proof that, under certain assumptions, a unique, stable, market-clearing equilibrium exists among rational economic agents has some clear explanatory and public policy relevance. One thing that is obvious about economies under all but the most extraordinary conditions is their stability: changes in the price of one commodity, even large swings in its price, do not result in the entire system of prices lurching about wildly or exploding into chaos. Market prices show a great deal of stability and apparent readjustment. This is a fact about the economy that needs explanation, and the proof that a competitive market among rational agents is stable is accepted by economists as providing one. The policy implication is more obvious. We know that the state central planning system of the Soviet Union and its client states broke down because it could not avoid shortages and surpluses. Economists have argued that the reason is that only a competitive market can minimize shortages and surpluses, and such a market is just what state central planning forbids.

The power of this proof—its application in explanation and policy, and the limitations under which it applies—raises many questions for the philosophy of social science, as we shall see. Meanwhile, the theoretical beauty of an explanation of how a self-regulating, welfare-optimizing institution like the competitive market emerges from the myopic self-centered, selfish, individual decisions of rational agents made rational choice theory an object of emulation throughout the social sciences. In political science, in sociology and anthropology, researchers began to treat prevailing institutions and practices as unintended and unforeseen equilibria that reflect the adding up of a large number of individual rational choices of shortsighted individuals who do not intend these welfare-producing outcomes but nevertheless benefit from them.

Once rational choice theory began to expand its influence in the social sciences during the last decades of the twentieth century, examples of such hidden-hand processes producing beneficial unintended outcomes began to be noticed throughout human affairs. Some obvious examples include the way drive-on-the-right traffic patterns emerged everywhere except the United Kingdom without anyone deciding on it. In the UK, of course, the spontaneous order just happened to go in the opposite direction. Word meanings and other aspects of language use are cases in point. Another example is the emergence of paper currency as a vehicle for exchange. If people began to really think about what a dollar bill or five-pound note is actually worth, paper money would lose its value and the benefit it confers on everyone would disappear. Its use is truly a case of spontaneous and unintended order. Then there is the fact that single-member plurality voting ("first past the post voting") produces stable political equilibria at outcomes close to the preferences of the median voter, even when none of those who win the elections and few of those who vote share these preferences. Indeed, one is tempted to say that such electoral systems persist because they have this function, unnoticed and unintended by any individual person who participates in these systems. Recall that this raises the problem of apparent teleology or purposiveness we identified in Chapter 5.

Social scientists, especially economists, who invoke such functions face a problem, one the economist Friedrich Hayek called the problem of *spontaneous order*: how can something come to exist with all the hugely beneficial features of the price system of an exchange market that remains close to some optimal equilibrium? Social scientists cannot call on God's benevolence as an explanation of its existence. Nor, as Hayek showed, could many of the social institutions that solve problems faced by human societies be the result of intentional design, creation, or maintenance by individuals or groups of them. No one can create or maintain a market price system. This problem—most acutely manifested by the price system, but widespread in human affairs—requires a scientifically acceptable answer. We return to it in Chapters 10 through 12.

Meanwhile, the potential explanatory and policy-relevant power of rational choice theory in the hands not just of economists, but of many other social scientists, makes it indispensible that we assess its scope and limits.

FOLK PSYCHOLOGY AND MARGINAL UTILITY

Marginalists not only held that utility is cardinally measurable, they also assumed that it is "interpersonally comparable." That is, it makes sense to say how much more one person desires something than another person desires it. These differences are measured in utilities that provide units for measuring the strength of desires. Marginalists were never able to show how we can interpersonally compare the utilities people derive from goods, but they assumed it was at least in principle possible. The technicalities were to be left to the psychologist.

If utilities come in definite units that can be compared between people, they would be important not only for their apparent explanatory power. They also enable the economist to make strong judgments about welfare and to advise on the distribution of benefits by the state. Marginal utility theory provided a strong foundation for claims about which distribution of benefits would do the most good in a society. If utilities are measurable and if the amount of utility each successive unit of some good provides—its marginal utility—declines, then $1,000 provides fewer units of utility to a millionaire than to someone at the poverty level. Thus, a government with $1,000 to spare, committed to maximizing total utility, would be justified in giving the money to the poorer person. The notion that $1,000 will provide greater satisfaction of wants to a poor person than to a rich one seems obvious—indeed, it is another piece of folk psychology. But it cannot be established to be true unless the difference it makes to people's levels of satisfaction is, at least in theory, measurable in cardinal units.

But no one has ever been able to identify the unit quantity of utility in the feelings of satisfaction, happiness, or pleasure that emerges in the brain when we secure anything we want. Perhaps neuroscience will one day do this, though in the last chapter we identified some of the problems facing the application of neuroscience in the study of human affairs.

But there is a much more immediate problem: the trouble with the claim that all economic agents are cardinal utility maximizers is that it just seems false. People frequently seem to do things that preclude the maximization of their utility. Consider acts of altruism, charity, or the frequent willingness to settle for good enough when the best may well be available. Of course, we can defend the theory by saying that these appearances are deceptive. Altruists really enjoy their good deeds, and philanthropists receive nonmonetary rewards—esteem, public notice, and the like. When we settle for less than the best or make a mistake, it is because the costs of finding the preferred alternative or of the correct calculations are too high. But these ploys open up marginal utility theory to the claim that it is unfalsifiable. The real nub of the problem is one we have faced before: we cannot really tell which—if either—of these criticisms is correct. Is the theory false or is it vacuously true? The reason is that there seems no way to measure cardinal utility. Therefore, there seems no way fairly to test the theory that people always maximize utility. Let us see why.

To begin with, we have here another case of a generalization that has to be employed to test itself, to establish the initial conditions to which it has to be applied. For the method marginalists offered to establish cardinal utilities already assumes that people are utility maximizers. We begin by giving an agent, say, ten apples. We then stipulate that the tenth apple the agent received provides one unit of utility. Now we offer the agent some orange juice in trade for apples. We observe how much orange juice he trades away the tenth apple for. This amount—say, five ounces—must then also be worth one unit of utility to the agent. By giving him the juice and repeating the experiment—trading orange juice for apples—we can determine how much juice is required to provide a second unit of utility. The marginalists held that marginal utility declines: more than five ounces of orange juice would be required to get the agent to give up the ninth apple. Marginalists often grounded this declining marginal utility in introspection: you can try the thought experiment on yourself to confirm it. But note that the experimental method of measuring utility will work only if the agent is rational, that is, maximizes utilities.

Moreover, the amount of utility a good produces depends on the availability of other goods. The utility of mustard for a person may be quite low in the absence of a hot dog, and the utility of a hot dog may be quite low in the absence of mustard. Together they each have a higher utility than separately. Mustard and a hot dog may have a higher utility still if beer is available, or if they are consumed at a baseball stadium . . . under sunny skies . . . with your favorite team playing . . . and winning . . . by a large margin . . . in the 8th . . .

The point is that we cannot undertake the experiment to measure the cardinal utility of each commodity by “pairwise” comparisons. The amount of utility each good generates will depend on a vast number of other goods present or absent. But that means that utility isn’t cardinally measurable the way mass is: utility is not a property things have independent of other things. Rather, it varies even when the good that provides the utility does not change. That makes the problem of measuring cardinal utilities difficult and the problem of actually employing them insurmountable. Since we don’t know what the other available goods are that we must hold constant when we measure the cardinal utility of one good, we can’t easily measure it at all. Since cardinal utility varies with the other goods available, knowing a good’s cardinal utility relative to one set of goods will not enable us to infer its utility relative to another set. If measurability and applicability are required for a legitimate scientific concept, cardinal utility is in serious difficulty: there is no natural zero amount of it; there are no ways to measure amounts of it; its amounts don’t remain constantly correlated with any observable states.

What is more, cardinal utility theory seemed far too psychological for economists. The notion of declining marginal utility and the appeal to introspection that often supported it were an embarrassment to rigorous economic analysis. After all, since there was no way of looking into the minds of consumers, the theory's basic claims were completely untestable. In the heyday of behaviorism in psychology, this was a serious embarrassment for economics.

Fortunately, by the early part of the century, mathematical economists were able to show that most of the important results of theoretical economics—including the laws of supply and demand—could be derived from a set of assumptions about rational choice free from almost all psychological content, assumptions that didn't require cardinal measurability, but only that utilities be rank-ordered instead of numerically weighted. This is the theory of ordinal utility or ordinal preference. It holds that:

1. For all possible pairs of commodities, the rational agent prefers one to the other or is strictly indifferent between them [the assumption of comparability].
2. For any three commodities, *a, b, c,* all rational agents who prefer *a* to *b,* and *b* to *c,* prefer *a* to *c* [the assumption of transitivity].
3. A rational agent chooses the available commodity that maximizes his preference.
4. Economic agents are rational—act in accordance with 1–3.

There is no assumption here that commodities provide numerical units of utility, only that there is a rank order of preference among all the goods or other alternatives available to the agent. Where the marginalists were committed to assigning consumers definite (though unknowable) amounts of utility from each alternative, ordinal utility theorists needed them only to rank available alternatives as most preferred, second most preferred, third most preferred, et cetera.

This more limited approach to rationality enabled economic theory to derive most of the same theoretical results (including the remarkable signaling power of a competitive market) that the cardinal theory provided, without the dubious excess baggage of cardinal utilities. But there was a catch, a price to be paid in the policy relevance of economics. Based on ordinal utility theory, economics had to forgo its claim to justify certain apparently attractive social welfare policies. For instance, economic theory could no longer sanction giving $1,000 to the poor person instead of the rich person on the grounds that the poor one would get more satisfaction from it. Ordinal utility does not allow for interpersonal comparisons the way that marginalism did. On this approach, we cannot say that one person prefers a commodity more than another one does, because we have no units in which to count the strength of either one's desire.

At most, ordinal utility theory allows us to make only a very weak claim about how to distribute goods in ways that attain beneficial results for the largest number of people: given a distribution of goods to people, ordinal utility theory tells us whether there is another distribution that makes at least one of the people better off without making anyone else worse off. For instance, if we give one banana to each child in the class and among them one child doesn't particularly like bananas, we can increase welfare in the group by giving that child's banana to someone else. But we cannot tell how much we have increased welfare, or to which child giving the banana will increase total welfare in the group the most. But whether there is a distribution in which no one can be made better off except by making at least one person worse off is a question that ordinal utility theory can empirically address: just vary the distribution among the people and ask whether any new distribution makes at least one person worse off. If the answer is yes, every new distribution does make at least one person worse off, then the original one was *Pareto optimal*. The distribution is called a Pareto optimum after the nineteenth-century Italian economist who first defined it. Pareto optima are pretty easy to attain, they are better than Pareto non-optima (where we can make some people better off without making anyone worse off), and they are not difficult to empirically establish. But the criterion of Pareto optimality that ordinal theory allows is far too weak to underwrite any very radical redistribution of income or wealth.

The result has been to remove modern Western mathematical economics from much of the debate about economic equality and exploitation. Many economists did not view this restriction on economic theory as a defect. They held that, like theory in natural science, economic theory should describe only the way things are. It should be value free and not take sides on normative issues. This is an important subject, one to which we shall return in Chapters 8, 13, and 14.

THE ECONOMIST AS BEHAVIORIST

Though it avoided the defects of cardinal utility, ordinal utility theory did not allay the nagging doubts about the falsity or vacuity of rational choice theory. After all, consider how this theory bids us measure the preferences of agents. We present a large number of different pairs of commodities to an agent, and on the basis of her choices, we construct a preference ranking. Of course we cannot present all possible pairs to the agent, but we don't need to. We can extrapolate pretty safely from the alternatives we actually give, if we are careful. This method will work provided that, first, the agent we are testing is rational and, second, her tastes do not change during the period of the test or before we apply the test results to predict her further choices.

But suppose a person behaves irrationally; suppose he violates the second condition above, the assumption of transitivity. People seem to do that frequently, and if they do, then the ordinal theory is disconfirmed, either because people are after all irrational or because even rational people make mistakes. Of course it is easy to insulate or protect the transitivity assumption of the theory against any counterevidence. Just claim that apparent violations are not real violations. Instead, insist that they are changes in taste. For instance, when I was a boy, I chose bubble gum over licorice, and both over peppermints. Now I choose peppermints in preference to licorice, and both over bubble gum. Yesterday I chose escargots over oysters; the day before, oysters over clams; today, I choose clams over escargots. My choice is not irrational; my tastes have changed. The trouble is that there is no way of distinguishing within economic theory between change in taste and irrationality. And there seems to be no way outside of the theory to tell change of taste from intransitivity, if economic theory is folk psychology formalized. For the way we actually tell when people have made a mistake as opposed to changing their tastes is by asking—by using [L].

The problem of distinguishing taste changes from violations of the principle of rationality is a rarefied theoretical one. But it reflects the fact that rational choice theory, for all its formalization, is no better at explaining and predicting the details of particular economic agents' choices than folk psychology is. This fact has led economists to fundamental reinterpretations of the aims and claims of economics. One reinterpretation excuses economics implicitly from the task of explaining individual human action. The other reinterpretation does so explicitly.

The implicit excuse is founded on a self-consciously behaviorist interpretation of rational choice, the "theory of revealed preference." The trouble with ordinal utility theory, some economists said, was that it was still too psychological. We do not wish to make any assumptions at all about what goes on in the heads of agents. We don't even want to use the notion of utility, no matter how minimally interpreted. And we don't need to. The only requirement the economist needs to make is that behavior is consistent, no matter its causes. These causes of consistency we leave to psychology. Consistent behavior means merely this: if an agent chooses *a* over *b* when both are available, then he does not choose *b* over *a* when both are available. On the basis of his choices, we can build up a preference map for the agent, but we need not attribute this preference map to him. All he does is make consistent choices that reveal a consistent pattern. The psychological machinery behind this consistent pattern is a matter of indifference to economics, and its theory of choice should be silent on the question. The agent's behavior is described as "revealed preference," but that is a misnomer because there is no commitment to a psychological preference revealed by choices that the preferences cause. There is just the consistency of choices, and that is all we require to derive all the standard results in the theory of consumer choice. The doctrine of revealed preference was said to free economics from the very notion of preference, from all dependence on the concept of utility, and from any psychological theory—folk or scientific.

It is indeed an interesting fact that almost all the results in economic theory that were once thought to require cardinal utility as a foundation, and then thought to require ordinal preference, turn out to be derivable from an extremely weak assumption that makes no claim whatsoever about the psychological causes of individual choice. This really is behaviorism of the most thoroughgoing sort. One important thing to note is that the adoption of revealed-preference theory limits the domain of economic theory even more than the surrender of cardinal utility does. Recall that the shift to ordinal utility excluded interpersonal comparisons and thus severely restricted the scope of economic judgments about welfare. With revealed preference, we surrender as any aim of economics the explanation of individual choice from the assumption that people are rational.

Revealed-preference theory in effect tells us that the starting point of economics is the consequences of, not the causes of, individual choice. To explain individual choice, we need to make some assumptions about what produces this behavior. But if economics refuses to assert the existence of a preference ordering that the individual actually has, independent of his behavior, then it cannot explain this behavior.

That may indeed be viewed by economists as a solution to the problem their theory shares with folk psychology: the problem of being unable to provide improvable explanations and predictions of individual action. The solution is that economics does not aim at such explanations and predictions. Its concern is only with the consequences of choice, not choice itself. Taken seriously, this attitude means that talk about preferences revealed in behavior is just a useful fiction, a handy instrument. It's just a convenient description of the behavior from which all results about markets and economies that interest economists follow. In some ways this approach makes the problem of spontaneous order, discussed above, even more serious. Recall that the original problem is to explain how the market, for example, arose and persists among people who do not aim at creating it. Now we have to add to the problem, explaining its persistence among people about whose beliefs and desires the economist has nothing to say at all.

Furthermore, it is easy to complain that changing the subject away from explaining individual choice does not provide a solution to the problem of explaining human action. Moreover, the new claim that agents are rational just in the sense that their choices are consistent does not really avoid the problem of falsity versus vacuity that haunts [L]. One reason for this is easy to see: consistency in choice among three goods implies transitivity, and we have seen that there is no way to distinguish violations of transitivity in choice from changes in taste. Revealed-preference theory has no room for the notion of taste—a psychological matter. But short of being disconfirmed every time a person's tastes do change, the transitivity assumption needs some qualification or ceteris paribus clause to remain plausible. The obvious qualification of assuming no changes in taste is not available to revealed-preference theory. For invoking changes of taste is a claim about people's psychological states after all. It is tantamount to surrendering the claim that revealed-preference theory is a purely behavioral doctrine, with no commitment to the psychological sources of consistency in choice behavior. Invoking tastes to protect transitivity readmits psychology—that is, desires—into economists' theory by the back door.

Furthermore, consistent choice behavior is rational only if, in addition to tastes remaining unchanged, beliefs do as well. If there are changes in my beliefs about the commodities among which I must choose, then sometimes the rational thing to do is to choose *b*, even though in the past I chose *a*. Of course, this possibility does not arise so long as we maintain economists' standard assumption of perfect information. Once this assumption is relaxed and beliefs must enter explicitly into determinants of choice, the whole pretense of the behaviorists of revealed-preference theory must be surrendered. For there is no way to read my beliefs off from my behavior except against some background assumptions about my preferences.

In their simplest models, economists try to hold beliefs constant so that they can ignore the impact of differences in beliefs on behavior, by an assumption that economic agents—consumers and producers—have "perfect information." They are fully acquainted with all matters they need to know about, that is, to believe, in making their choices. But often the assumption of perfect information must be surrendered if the theory is to be applied to the real world. And when economists do this, psychology enters economics again, this time through the front door.

The most fertile and influential theory of economic choice under conditions of uncertainty, the theory of "expected utility," originated by John Von Neumann and Oskar Morgenstern, involves this very triangular relationship among beliefs, desires, and actions. The central role of the theory of expected utility in contemporary economics has reinforced both the discipline's commitment to explaining individual action and economic theory's character as a formalization of folk psychology.

The theory of expected utility begins by resurrecting cardinal utilities, though not interpersonally comparable ones. It does so by a subtle use of the psychological assumption that people are utility maximizers. The method works like this: we offer a rational agent choices between, say, a 100 percent chance of $100 and a lottery ticket providing an 80 percent chance of $200 and a 20 percent chance of nothing. If the agent chooses the certainty of $100, its utility must be greater than that of the lottery ticket. If he is indifferent between $100 and the ticket, then their utilities must be the same. Assume he is indifferent. Therefore, an 80 percent chance of $200 worth of utility plus a 20 percent chance of nothing equals a 100 percent chance of $100 worth of utility. That is, .80 × (utility of $200) + .20 × (utility of $0) = 1.00 × (utility of $100). Since 1.0 divided by .8 equals 1.25, by simple rearrangement, the utility of $200 = 125 percent of the utility of $100 to our agent.

So, if we stipulate that a 100 percent chance of $100 provides our subject with one unit of utility, then we can determine the amount of utility that any other commodity will provide him. We simply offer him choices between $100 and lottery tickets in which we vary the probabilities of getting the other commodity or nothing, until he says he's indifferent. Then we calculate what the utility of the commodity must be to make him indifferent between the $100 and the lottery ticket.

Once we have identified the cardinal utilities of the agent by these means, we can explain and predict choice under conditions of uncertainty. By combining his cardinal utilities and his incomplete information—his probability beliefs—we can derive the rational agent's utility-maximizing actions. But to apply this recipe to actual choice under uncertainty, we need to establish what a person's beliefs about the probabilities are. How do we do that? We use the same lottery method we employ to measure his utilities. Only now, given his cardinal utilities for commodities, we offer him choices between certain outcomes and lottery tickets for known utility values with the probabilities left blank. Whenever he expresses indifference between the certainty and the lottery ticket, we can use the same formula, with the utilities as data and the probabilities as unknowns, to calculate the probabilities he attaches to future available outcomes, that is, his beliefs about how likely they are to be available. Thus we can either work back from choices and beliefs about probabilities to desires—utilities—or from choices and desires, expressed as utilities, to beliefs about probabilities. This theory of choice under conditions of risk is truly folk psychology formalized to a very high degree.

But we cannot adopt expected-utility theory as a refinement of the economist's theory of rational choice if we take seriously a behaviorist's interpretation of economics that simply reads individual choice out of its domain altogether. Since behaviorism about rational choice does not seem to solve its problems in any case, the exclusion of individual choice is not much of a reason to forgo expected-utility theory. Moreover, the latter's prospect of accommodating economic theory to the fact that people do not ever have the sort of complete information usually assumed by economists is in itself a strong reason to embrace the theory of expected utility. Additionally, its employment by political scientists, operations researchers, and students of management seems to have improved our abilities to explain (though not predict) some aspects of behavior beyond the powers of unformalized folk psychology.

However, it is evident that people do not seem to act in strict accordance with the theory of expected utility any more than they act in accordance with revealed-preference theory. Any claims expected-utility theory makes to have increased explanatory unification and predictive precision are controversial. After a certain point there will be serious limits on its further improvement, because of the impossibility of measuring utilities and probabilities independently of assuming the truth of expected-utility theory and the truth of [L]. And at that point it will face the same problems that bedevil folk psychology.

INSTRUMENTALISM AND MODELING IN ECONOMICS

Difficulties in the employment of rational choice theory to explain and predict the behavior of individual economic agents have led economists to adopt a strategy for reading individual choice out of the domain of economic theory. Like other aspects of economic method, this one, too, has found favor among noneconomists who have adapted rational choice theory to their own subjects.

Rational choice theory is thus to be viewed as a calculating device, a convenient model that helps us systematize our expectations about markets, industries, and economies. Its truth or falsity as a set of claims about what makes individual agents tick is irrelevant to the intended domain of the theory. Of course the theory makes many contrary-to-fact assumptions: that individuals have a complete transitive preference order that maximizes expected utility; that they have complete information; and other assumptions we will identify in this section. Some economists argue that doubts about the truth of these highly idealized assumptions about individuals are simply misplaced. A theory is to be judged not on the truth of its assumptions, but on the confirmation of its predictions for observation. That is because predictive success is the sole mark of scientific success, at least in economics.

The strategy is a variant of one well known in the philosophy of science: instrumentalism. According to this doctrine, scientific theories should not be treated as literally true or false claims about the world. Instead they are devices for systematizing our observations. The claims of scientific theories about unobservable entities and forces that underlie observable phenomena may be viewed as heuristic devices, useful fictions, that help us predict observations, but not as referring to existing things and processes. A somewhat weaker version of this approach holds, not that theories work by postulating fictions, but that we can never know whether their claims about unobservable reality are correct, since our knowledge extends only as far as observation. Therefore, we should be agnostic about the theoretical claims of science, merely using the theories that work, without committing ourselves to their truth. It should be evident that instrumentalism has provided much of the motivation for behaviorism, both in experimental psychology and in economics. It lets us use words like *belief*, *desire*, *expectation*, *preference*, *information*, *uncertainty*, without having to take them seriously as naming mental states of consumers or producers.

There seems much to be said for this approach to rational choice theory. Of course, not all agents are rational all the time or, perhaps, even much of the time. But from the unrealistic, idealized assumption that people are rational, many important large-scale economic phenomena, such as the equilibrium price level, can be derived, and thus predicted. So the argument goes.

We must distinguish this view from an alternative that holds that although no one behaves rationally all the time, the individual deviations from rationality are distributed along a bell-shaped curve in such a way as to cancel each other out. Thus the average of all individuals' behavior falls close enough to rationality for it to explain and predict aggregate economic phenomena. If this view is right, there is no special mystery about why rational choice theory is explanatory "in the large," even though it is false or quite weak in its explanations and predictions of individual behavior.

But if, when aggregated, approximate rationality were the case, or equivalently if on average people are rational, we would naturally seek an explanation of why individual deviations are so conveniently distributed. The distribution cannot be accidental, and one suspects that the individual divergences should be explainable by appeal to a finite number of different interfering forces, deflecting individuals to greater or lesser degrees away from the optimally rational choice. If we could find these factors, we could add qualifications to our theory of rational choice that would enable us to improve our explanations and predictions of individual behavior. We could measure the degree to which the interfering forces were operating in any given case. This defense of rational choice theory is far from an instrumentalist one. And if it could be substantiated, it would establish the credentials not only of the economic theory of rational choice, but also of the whole strategy of providing intentional explanations of the large-scale consequences of aggregated choices. Of course, in Chapters 4 and 5 we repeatedly ran up against obstacles to identifying the disturbing forces that infect the measurement of belief and desire and that block attempts to improve such theories.

If economic agents' behavior simply deviated from the rational in a regular and replicable way, then there wouldn't be any obstacle to interpreting rational choice theory as a set of statistical regularities. The instrumental interpretation of the theory would be gratuitous. But the instrumentalist wants a justification for using rational choice theory, even where there is no evidence that individual behavior is on average rational in the way that would explain large-scale economic regularities.

Thus, the instrumentalist claims that the only basis on which to assess the rational choice assumptions of economic theory are their consequences for observations of aggregate phenomena that actually concern economists: are the predictions of the rational choice theory about the effects of a rise in the price of oil on aggregate demand for oil borne out by evidence? Are its predictions about how the stock market responds to a change in the interest rate or in the money supply confirmed? If we employ the theory to predict the effect of an excise tax or a monopoly and to design public policy that turns on such predictions, will our policy goals be attained? These are the sorts of questions that concern economists. If the theory of rational choice gives us the right answers about these broad-scale economic questions, then, the instrumentalist argues, it is all the justification it needs. The theory will have proved itself a reliable tool in the service of economists' scientific aims.

If, as some economists write, the theory of individual rational behavior is just a stepping-stone to economists' real interest in groups of individuals, then perhaps the best view to take of the theory is that of a convenient fiction or a calculating device, a model not intended to describe anything accurately. Assuming that economic agents behave as required by the three axioms of ordinal-utility theory is just a convenience for organizing our analysis of large-scale economic processes. It is one we combine with further assumptions to build models of the economy that are not to be assessed for their truth, realism, or descriptive completeness. They are tools that provide output, in the form of predictions about market processes, for input, in the form of data about initial conditions obtaining in the market.

This view of theory as a "black box"—a useful instrument, and not an attempt to give a true description of economic reality—often goes together with an insistence that in economics (and social science inspired by it) what is sought are *models*, often mathematical or quantitative models. These are intentionally idealized, highly incomplete sets of assumptions, sometimes even intentional caricatures that no one supposes describe economic processes accurately, but are for that reason particularly useful. The expression *black box* suggests that we don't know or care what goes on "inside the theory"; it is just a device for generating outputs given inputs. In this case, if we input people's preference rankings and the available inputs to production, the black box is supposed to automatically spit out the list of outputs that will make people as well off as we can provably make them.

The theorems of general equilibrium described in the first section of this chapter are good examples of such an idealized model and its various uses, ones that do not require it to be a true description of reality. This model incorporates the model of rationality and adds several other plainly idealized assumptions to derive some important conclusions about the economy as a whole:

1. Agents are rational; they satisfy the axioms about ordinal utility given above.
2. Agents have complete information about all alternatives open to them in production and consumption; no one has "inside information" that others lack.
3. All commodities are infinitely divisible: just as one can buy or sell any amount of water or wine, one can purchase or sell any amount of a car—the whole car or half of it or three-seventeenths of it.
4. There are constant returns to scale in production—building a bigger factory or adding more workers won't increase or decrease the efficiency of production.
5. It is possible to purchase or sell, today and every day into the future, any good or service for future delivery at any time and place.

With these five assumptions, it can be proved mathematically that:

> There is a unique, stable, market-clearing equilibrium in every market for every commodity in this economy, and no one can be made better off without making at least one person worse off.

The idealizations involved in these assumptions make the economy they describe a model. Now, why should such a model be of use to economists? There are several alternate answers to this question, only one of which vindicates an instrumentalist approach to models and theories in the social sciences.

One account of the role of models in economics and elsewhere is that despite their idealizations, they are close enough to the truth to provide at least tentative explanations that will be improved by further refinement. We'll return to this idea. Another argument for modeling is that the model identifies and highlights important aspects of economic phenomena that might otherwise be hard to notice in reality, or enables us to trace relationships that are not obvious in the data itself.

A related argument is particularly important when it comes to public policy applications of economics. The model of a competitive equilibrium can be proved to have very desirable consequences for the welfare, happiness, and preference maximization of consumers and producers—no one can be made better off in such an equilibrium without making at least one person worse off. So, the model thus gives us guidance about how to approach this desirable outcome in the real world: adopt policies that to the extent possible increase the match between reality and the model's assumptions. For example, break up large companies likely to exploit increasing returns to scale to become monopolists that destroy the competitive market, provide reliable futures markets for commodities, effectively enforce prohibitions against insider trading and other failures of the "complete information" assumption. Or force people to act more rationally. Economic models can be tools in policy guidance even when they are radically idealized, very unrealistic, and quite simple in their structure.

But let's return to the instrumentalist justification of models as black boxes, heuristic devices, tools that meet our needs to organize economic and other kinds of aggregate data. The instrumentalist argues that a more realistic theory, studded with qualifications, ceteris paribus clauses, introducing more and more of the factors that obtain in the real world, would be impossibly complicated to employ, would provide no clear-cut predictions, and would fail to connect and unify diverse phenomena into a manageable system. Thus, unrealistic models are essential to economics, as they are to the rest of science. And the most important of these is the theory of rational choice. Questions about its truth or its explanatory and predictive power for individual action are beside the point as far as economics is concerned. This conclusion will be seconded by some political scientists, some sociologists, and even some historians, those who focus on the behavior of groups and attempt to explain their behavior by appealing to this theory.

The adequacy of this view hinges on many issues: one is the philosophical issue of whether instrumentalism is a tenable approach to the nature of scientific theories. This question will give little pause to working economists. They are no more interested in philosophy than experimental psychologists usually are—in fact, less! A more immediate question is whether economic theory has actually been as successful in systematizing and predicting aggregate economic data as this argument requires. For unless the economist's predictions about aggregate economic phenomena are well confirmed or at least improving, the claims on behalf of his black box, the theory of rational choice, will be moot.

Before an unrealistic theory can be accepted on instrumental grounds, it must be shown to actually be a good tool, to do things that other theories cannot do, to improve the accuracy of our expectations about the future, given information about the past and present. Has the theory of rational choice met this instrumental test? That is not a question on which a text in the philosophy of social science can take sides. It is a factual matter to be decided largely by economists. But the accuracy of economic predictions employing the theory of rational choice is a necessary condition for the cogency of the instrumentalist defense. In what follows, let's assume the condition is met. The result will in fact be an argument against an instrumental reading of rational choice theory.

The most disturbing issue such an approach to economic theory faces is the question of why the theory of rational choice and the rest of the idealized and unrealistic assumptions are so useful in systematizing and predicting aggregate economic phenomena.

We may illustrate the problem by appeal to a simple example from natural science. Consider the ideal gas law, $PV = nRT$, according to which the temperature of a gas is a function of the product of its pressure and volume (where R is a constant of proportionality). There is a model in the kinetic theory of gases that systematizes this regularity along with several other generalizations of thermodynamics. It is the well-known "billiard-ball" model of a gas: gas molecules are assumed to behave like billiard balls on a table. Like such balls, the molecules obey Newton's laws of motion, and the aggregate values of their individual mechanical properties are identical to the thermodynamic properties of the gas as a whole. Thus, the temperature of the gas is equal to the average kinetic energy of the molecules it contains.

But kinetic theory is highly unrealistic. For PV = nRT follows from the model only on two highly unrealistic assumptions: that molecules are point masses, that is, they have mass, but no volume; and that there are no intermolecular forces acting between the molecules—they just bounce off each other with perfect elasticity.

No one will reject the kinetic theory of gases just because it embodies unrealistic assumptions. Thus, the example seems to be an analogical argument in favor of retaining other theories that are predictively successful even though their assumptions are unrealistic. But, we must ask ourselves, why does the billiard-ball model function as a good instrument for accommodating the behavior of gases? Obviously, because the two unrealistic, false assumptions are pretty close to the truth. Evidence for that comes with improvements in the data of thermodynamics and improvements in kinetic theory. Thus PV = nRT seems to hold for moderate values of pressure, temperature, and volume; at great pressure, low volumes, and very high temperature, it is strongly disconfirmed: gases no longer obey the formula. But by adding assumptions to kinetic theory about the small but finite volume of molecules and the infinitesimal intermolecular forces that actually obtain, we can derive an improved and only slightly more complicated version of PV = nRT that does accommodate this new data. This new, more realistic model explains why the old, simple one worked so well. In fact, it constitutes an argument for continuing to use the older, less realistic model out of convenience when dealing with gases at moderate values of *P*, *V*, and *T*.

Therefore, the explanation of why the billiard-ball model is a good instrument is that it is pretty close to being true and that there is another model remarkably like it, but even closer to being true; that fact explains why the former is a good instrument. That should be no surprise, for nothing is a good instrument by accident. For every good instrument, whether a hammer, a computer, a linear accelerator, or a theory, there must be an explanation of why it works so well. The simplest explanation in the case of a theory is that it is true or close enough to the truth to be relied upon.

This is not an explanation the instrumentalist about rational choice theory can accept. For the initial motivation of an instrumentalist interpretation of rational choice theory is to be able to employ it regardless of whether it is true or close to true.

Of course, following the advice of instrumentalist philosophers, economists can block this argument by refusing to answer the question of why their black box works well. This question, they insist, is not one that the science of economics—or any science—has to answer. For answering it does not increase the predictive power of the theory, and in the end that is the sole goal of science.

This assertion is a distinctively philosophical one, just the sort of claim many economists wish to avoid. Near the end of Chapter 4 we came to the conclusion that the social scientist wedded to intentional explanation must eventually confront the fundamental philosophical problem of the relation of the mind and the body—as reflected in the problem of intensionality versus extensionality. Psychologists' and economists' attempts to circumvent that problem bring them face to face with another one—this time a fundamental epistemological problem. For their assertion that science has but one goal, and that predictive success is that goal, is not to be settled by factual findings in any of the individual sciences. The question, What is the goal of science? ultimately comes down to what counts as knowledge.

Most people agree that knowledge is the ultimate goal of science. Perhaps the only ones who don't agree are the instrumentalists who claim predictive success is its sole goal. And even some of those instrumentalists make common cause with philosophers and social scientists who hold that predictive success is at least a necessary means of certifying scientific knowledge, not a substitute for it. By contrast, when social scientists and philosophers hold out another goal, such as intelligibility, as the one we should pursue, they implicitly or explicitly endorse a different epistemology. Ultimately the choice between naturalistic social science and interpretative social science comes down to a decision about epistemology, as we shall see by the time we have finished the next several chapters.

THE ECLIPSE OF BEHAVIORISM IN PSYCHOLOGY AND ECONOMICS

Behaviorism's attempt to deal with the problems of an intentional social science did not succeed. The problems facing naturalism could not really be circumvented. In the next few chapters we examine the way interpretational or antinaturalistic approaches to social science have dealt with the same problems. But the reader must not be left with the impression that behaviorism proved a fruitless dead end. Because it dealt with the limitations of behaviorism, psychology—and especially what came to be called cognitive science—emerged ready to face squarely the problems of intentionality. Ironically, it did so in some measure by filling the vacuum that economics created when it turned its back on individual choice. Cognitive science begins very much with experiments undertaken to reveal how individuals make economic choices under laboratory settings, and how these choices—behaviors—can be systematically affected by varying the conditions under which they are made. But the explanation of these systematic differences in behavior could only be provided by hypotheses about expectations and preferences, that is, desires and beliefs. The result was the explicit surrender of behaviorism as a philosophy of experimental psychology. It should not be surprising that such advances in psychology would eventually have an impact among economists, very much changing their attitude to both the relevance of psychology to economics and its prospects for improving prediction by increasing the psychological realism of its assumptions.

The strategy employed in the research program of cognitive social psychology is to subject a significant number of individuals to a choice problem, in which rational choice theory dictates one choice as utility maximizing, and the psychologist then observes whether the actual choices diverge from the rational choice in systematic ways. For example, in an experimental setting, subjects—often university students—may be given an amount of money and then invited to use some of it to bid for a commodity in an auction—something as simple as a coffee mug. In such a case the highest bid a subject makes should reveal the contribution that owning the coffee mug makes to his or her welfare, happiness, satisfaction, or utility maximization. Now, in a second experiment, the same subjects are given free coffee mugs and invited to auction the mugs off. In many replications of this sort of experiment on large numbers of people for a variety of different objects or amounts of money, the results are surprising, at least on the assumptions of rational choice theory. When selling the mug they got free, the subjects demand a significantly higher price for the mugs than their highest bids for the same mug in the first experiment. What the experiment shows is that ownership of a commodity interferes systematically with the rational valuation of its utility. In general, such experiments have confirmed the generalization that agents suffer greater utility declines from the loss of a commodity they already own than utility gains from acquiring the very same commodity. In ordinary terms, other things being equal, people typically dislike losing something more than they like gaining it. This explains why people won't bother purchasing replacement tickets when they have lost some, even though the value to them of the event the tickets are for is greater than the price of the tickets.

There are a large number of other such generalizations that have been uncovered by cognitive science about systematic departures from purely rational choice. In a similar experiment, subjects are given a choice between a small amount of money and an object, say, a pen generally known to be worth more than the money. In such an experiment, some percentage of subjects choose the pen. When subjects are offered the same choices, but an inferior third option is provided, say, a cheap pen worth less than the money, the proportion of subjects choosing the more expensive pen increases. Rational choice theory requires that the addition of an option less attractive than either of the original ones should have no effect on choice. Yet it systematically does. Over the past three decades, cognitive scientists, and now, increasingly, economists, have conducted experiments that have revealed such systematic departures from what rational choice theory predicts when it is treated as a theory of individual choice. The Nobel Prize in economics for 2002 was awarded to a cognitive scientist and an experimental economist for their contributions to this research program.

The immediate implications of this research for economics and psychology and for their philosophies and methodologies are quite different. The behaviors elicited in these experiments lead the psychologist to take seriously the causal significance of internal mental states like expectations and preferences, and to design further experiments to uncover the conditions under which they appear to depart from perfectly rational expectations and preferences, and the amounts by which they do so. In this research the controlled experiment employing large numbers of individuals subject to radically different stimuli is essential to ensure that departures from rationality are common and in the same direction for most people. When a significant departure from the prediction of rational choice theory is uncovered, the starkness of the choices and the large numbers of subjects provide a way of "canceling out" differences in individuals' beliefs that might rationalize the apparently irrational choice in individual cases. The generalizations that cognitive psychology can hope to uncover by such experiments will be statistical and not subject to much refinement just because of problems of intentionality. Transcending these limits will require the resources of neuroscience in the study of individual differences between people. Such studies may eventually enable us to explain and predict human behavior with improving accuracy, but only for one person at a time.

Among economists, this research spawned a school of thought that labeled itself *behavioral economics*, reflecting its acceptance of the findings of cognitive scientists about how behavior systematically departs from what rational choice theory dictates. For example, experimental research about differences in the way subjects value coffee mugs, depending on whether they bid for them or offer them at auction, can be used to explain differences in economic agents' willingness to sell a stock at a given price. Experimental psychology reveals that whether the price represents a gain or a loss is the real issue. Given the way people behave in the coffee mug experiments, psychologists are not surprised to see the following pattern among investors: Two otherwise similar investors both own stock worth $10 a share. One purchased it when it was $20 a share and the other at $5 a share. The investor who purchased at the lower price is much more likely to sell than the investor who purchased at the higher price. This fact is anomalous in the light of rational choice theory. Agents' expectations about the future value of the stock is all that rational choice theory would allow to be causally relevant to their choices! The immediate result for economics of taking on the discovery that choices are irrational in systematic ways is of course to enable it to explain the predictive failures of rational choice theory. The medium-term results we might expect are improvements in the realism of the theory's assumptions that will enhance its predictive powers for markets, industries, and other economic aggregates—just what the original behavioral/instrumental interpretation of economics denied was necessary or feasible.

Another area in which increased realism about the mind has had payoffs for economic understanding is the role of differences in knowledge—that is, well-justified beliefs—between buyers and sellers. A famous example, due to George Akerlof, may help. Recall the assumption of complete information required to derive the existence of an efficient market-clearing equilibrium. Now, consider a simple model of the used-car market. In this model used-car market, sellers know the real condition of their cars. They know in many cases that the car is in good condition. Buyers do not know this, and what is worse, buyers have some reason to disbelieve the assurances of sellers about the condition of the used car. In any case, buyers cannot verify sellers' claims about parts of the car by observation or experiment—that is, a searching examination by an objective mechanic. From these assumptions, we can derive the result that buyers will bid less than the full value of reliable used cars and that sellers will have to accept these lower prices, since they cannot assure buyers of their cars' reliability. The result will be a suboptimal equilibrium in which the supply of reliable used cars will be too low (since the price is too low), and thus sellers and buyers will be worse off than under complete information. This model will not predict the emergence of reliable used-car businesses such as CarMax in the United States, which warrantees its vehicles. But it certainly helps us understand their success in retrospect. For our purposes it is worth noting that it is difficult for an instrumentalist about economics or behaviorist about the mind to take information seriously, if it is to be found inside the head!

In Chapter 11 we will return to the problem of spontaneous order in economic processes and in social institutions. We will see there how approaches to choice that repudiate realism can actually help solve this outstanding problem raised by Adam Smith's powerful notion of an invisible hand.

Introduction to the Literature

Any textbook of microeconomics provides a useful introduction to rational choice theory, under the label the "theory of consumer behavior." An excellent history of the changes in utility theory is M. Blaug, *Economic Theory in Retrospect.* For an informal introduction to the theory of expected utility, consult B. Skyrms, *Choice and Chance.* An exposition of the theory with applications to political science is W. Riker and P. Ordeshook, *Introduction to Positive Political Theory.* The thesis that rational choice theory can explain everything of interest about human behavior is defended and illustrated in Gary Becker, *The Economic Approach to Human Behavior.* Becker's theory involves a novel interpretation of the conventional theory, an interpretation that he claims circumvents conventional criticisms of it, including those treated in this chapter. A. Rosenberg, *Economics—Mathematical Politics or Science of Diminishing Returns?* expands on these criticisms and also assesses Becker's "new" theory. An extract from Becker together with other important papers on rational choice theory, including Herbert Simon's notion of satisficing or "bounded rationality," can be found in J. Elster, ed., *Rational Choice.* H. Rachlin, "Maximization Theory in Behavioral Psychology," not only explicitly applies the theory of rational choice to animal behavior but also shows the close similarity between revealed-preference theory and operant behaviorism. D. Papineau, *For Science in the Social Sciences*, provides a philosophically sophisticated defense of rational choice theory as embodying laws of human behavior.

Instrumentalists in economics and others who wish to insulate the discipline from methodological scrutiny appeal to a famous paper by Milton Friedman, "The Methodology of Positive Economics," widely reprinted and available in Ryan's anthology, *The Philosophy of Social Explanation.* This anthology also contains an important criticism of Friedman's views by Ernest Nagel. The anthology by D. Hausman, *The Philosophy of Economics*, also reprints Friedman's paper, together with other influential documents on the nature of economics. Hausman's introduction to the subject is especially helpful. M. Blaug, *The Methodology of Economics*, gives a convenient history of the controversy that followed Friedman's paper. Both Blaug's book and Hausman's edited volume contain invaluable bibliographies. A relatively advanced but important discussion of rational choice theory is to be found in two papers by A. Sen, "Rational Fools" and "Behavior and the Concept of Preference," both reprinted in his *Choice, Welfare and Measurement.* The latter paper is also to be found in J. Elster's anthology.

An excellent introduction to the increasing importance of cognitive science to economics can be found in Matthew Rabin, "Psychology and Economics."

A classic and accessible example of how economists have begun taking information seriously is George Akerlof, "The Market for Lemons," which explains why a used car is usually priced below its value to the buyer.

Steel and Guala's *The Philosophy of Social Science Reader* includes several articles of considerable import in the debate about the aims and limits of economics: Harsanyi's "Advances in Understanding Rational Behavior," one of Daniel Kahneman's important contributions to behavioral economics, "Maps of Bounded Rationality: Psychology for Behavioral Economics," and Philip Petit, "The Virtual Reality of *Homo Economicus*."

CHAPTER SEVEN

Social Psychology and the Construction of Society

Interpretationalists have no interest in surrendering the belief/desire model of explanation. They are required to provide an account of exactly how such explanations work. They do so by invoking and analyzing the concepts of rule, norm, and practice that connect individual intentional states to actions and through them to social institutions. Interpretational social scientists insist that this approach to explanation reveals its foundation in language and its implications for the conventional character of a great deal of human life.

SOCIAL PSYCHOLOGY AND THE CONSTRUCTION OF SOCIETY

As a predictively improvable theory of human action, folk psychology leaves much to be desired. Behaviorist attempts to circumvent its weaknesses seem not to do much better. Nor do approaches that reject intentionality altogether for some allegedly more experimental approach seem to hold out immediate hope of greater predictive power. It is, of course, possible that human actions are fundamentally undetermined by causal factors. For all we know, people have free will—their intentions are beyond the reach of any predictive theory because they are uncaused.

Opponents of a scientific approach to human action and its consequences have occasionally employed free will as a premise in their arguments that no causal laws about human action are forthcoming. For example, some argue that since we have free will, human actions do not fall under any very strict regularities, known or unknown. But such regularities are what causation consists in, and what scientific explanation requires. Therefore, no scientific theory of human action is possible.

However, most social scientists and philosophers consider the doctrine of free will too controversial to figure as the starting place in an argument against a science of human action. They recognize that there may be philosophically far fewer controversial explanations of why we cannot improve on folk psychology's predictions. Some of these critics of a predictive science of human action may be motivated to find arguments against a causal science of human behavior by the conviction that we have free will and/or by the fear that such a science would threaten that conviction. But being motivated to find an argument is not the same as actually finding one. An argument that we ought not to seek a predictive social science starts by identifying a different goal for the human sciences. Instead of prediction, they should aim at understanding and achieve it by uncovering meanings.

One such argument begins with this observation: even if our actions were causally undetermined, our intentional explanations of our actions would continue to be accepted as illuminating them. That suggests that such explanations are not causal claims at all. If so, we should not treat folk psychology as a causal theory or the forerunner to one, to be improved upon by the employment of experimental methods. We need to view intentional explanations in quite a different light. If we can find the right way to understand the theory, perhaps the problems and puzzles that burden the causal interpretation of folk psychology will disappear. They will turn out to have been pseudoproblems, generated by the mistaken presupposition that in social science our aim is causal knowledge or predictive improvement.

Here the reader may recall the interpretationalist position set out in Chapter 2, "Rejecting Empiricism for Intelligibility," which constitutes a preamble to the antinaturalistic argument this chapter will examine. First we sketch out the approach to explaining human action in terms of its meaning, then we examine notions like rules and norms crucial to this enterprise. Finally, we consider, from the interpretationalists' perspective, whether this explanatory strategy can be reconciled with a causal treatment of desires and beliefs. In this next chapter we turn to how the interpretationalist parlays the importance of meanings into an entire theory, not just about social science but about society itself. At the end of this chapter, we will see how wide the gulf is between naturalism and interpretative social science.

THE HERMENEUTICS OF HUMAN ACTION

Human action is explained by interpreting it, that is, by giving it meaning or significance. That is not a new thesis; indeed, Plato argues explicitly in the *Phaedo* (99 a–b) that human action can only be so understood. *Hermeneutics*—a term that originally designated a branch of theology devoted to the exegesis of the Bible—has in the past century or so been pressed into service as a name for the science of interpretation in general. Students of hermeneutics insist that explaining action is a matter of meaning. It follows, therefore, that the methods of social science must reflect the influence of the distinctive human capacity for language and the learning of it. And the purely physical character of human behavior, whether captured in causal regularities or not, must be relegated to subsidiary importance in the social sciences.

Desires and beliefs explain action by making it intelligible, revealing its meaning or significance. Thus, we often phrase a request for explanation of an action in the words, "What is the meaning of this?" Of course we also ask about the meaning of natural events: "Does that ominous cloud in the west mean rain?" But this usage is metaphorical. Merely physical events do not have meaning by themselves. When it comes to human ones, matters are different. Hermeneutics takes the appeal to meanings quite literally. Finding the meaning of an action is equivalent to deciphering a text. Deciphering a text requires that we understand the language in which it was written. The language in which it was written consists in a series of rules. If the text is a poem, then we need also to identify the rules that govern its form—blank verse or rhyme, Italian or English sonnet, sacred or profane, and so forth. Once we have learned all the rules that govern the text, we know its meaning. There remains but little that the specific intentions of the author can add to our understanding of the text. Why the work was produced on one day rather than another may be of interest to the biographer and literary historian, but knowing the rules that govern the poem's meaning cannot give us this information. If our aims, however, are not to predict when a poem will be produced but to understand it once it has been written, this information may be of passing interest only. All the action is in discovering the rules.

Social science seeks the meaning of actions and events composed of them. Thus it needs to identify the rules that give these things their meanings. Unlike the rules of grammar, these rules, or *norms*, are usually unwritten, implicit, complex, and hard to state fully. Uncovering them and the meanings they convey is far closer to an activity like literary criticism than it is to natural science. The role of meaning in understanding human affairs explains why desires and beliefs are what we seek when we set out to account for actions. It explains our unswerving commitment to folk psychology and to its immunity from "scientific" criticism or "improvement." Recall the logical connection argument of Chapter 4. Desires, beliefs, and actions are logically connected in virtue of their intentional content, for the content of each implicates the others. That does not preclude the existence of a causal connection among them, as we saw. But it makes clear that the connection among them hinges on the sentences they contain and therefore on the languages in which those sentences are expressed. The causal connection, if any, among them are sidelights to the intelligibility that such linguistic connections provide.

In fact, according to some philosophers, the very notion of linguistic meaning must be understood in terms of desires, beliefs, and actions. The idea that actions are explained by giving their meaning thus reflects a basic fact about meaning in general. Therefore, far from being metaphorical, intentional explanations are the basis of linguistic meaning. When I nod my head in answer to your question, "Are you thirsty?" why does that mean, "Yes, I am thirsty"? First, because I want you to believe that I am. But I could have gotten you to believe that by making you notice the sweat on my brow, my parched lips, and my swollen tongue. And those phenomena don't mean, "I am thirsty"; they are just indications of thirst, evidence for my being thirsty. So meaning requires more than indicators. That's why dark clouds don't really, literally, mean it will rain. The nodding means "I'm thirsty" because it got you to believe I am, and because it got you to believe that I wanted you to believe I was thirsty, and finally, because I wanted the nodding to get you to have both beliefs: that I was thirsty and that I wanted you to believe I was. It's hard to keep these nested beliefs and wants straight, but each part is necessary to distinguish linguistically meaningful action from symptoms, signs, and indications of what we believe and want.

This analysis, however, leaves something crucial out. Why did I choose nodding my head instead of, say, shaking it to express assent to your question? Because I believed, correctly, that there is a *rule* in our language: expressions of assent are given by nodding; expressions of dissent are given by shaking. And though the story about the nested beliefs and desires was crucial to my actions' having the linguistic meaning of assent, the most significant factor in my nodding was my recognition of this rule. Like all the rules of language, this one reflects a fact about my linguistic community, a fact that anyone who hopes to understand my language needs to learn.

Here then is the key to understanding human action, that is, meaningful behavior. We need to identify the rules and norms under which it falls because they are what give it meaning. The rules under which actions fall are reflected in the intensional content of the desires and beliefs that lead to them. That is why desires and beliefs explain action. Human action is thus a matter of following rules, and honoring norms, and the aim of social science is to uncover these rules and norms.

Written rules and unwritten norms are facts about communities. Linguistic rules are facts about members of a linguistic community; chess rules are facts about the community of chess players. Traffic rules are facts about drivers, pedestrians, police officers, judges, and so on. Rules range from the obvious, such as drive on the right, to the esoteric, such as that paintings of the Annunciation of the Virgin always show her on the right. They range from the specific, like "*i* before *e* except after *c*, or when sounded like *a* as in *neighbor* and *weigh*," to the general, like [L]: "If you want *d* and believe that action *a* is one means to attain *d*, then do action *a*." Often rules and norms operate in mores and practices that are hard to articulate, even by those whose actions they govern. Like complex rules of grammar, we can't state them even though we can unfailingly detect violations of them.

To the extent that [L] governs behavior, it does so as a rule, not a law. How do we know? Like all rules, [L] has two features that no causal law has. First, we can break it; and second, breaking it is punishable. These two properties of rules are related, of course. For if we could not break a rule, there would be no need of a punishment to encourage compliance. In the case of many rules, the punishment is obvious. Break the traffic rules, and you are liable to fines—or worse. Break the rules of bridge, and you will lose points. Break the marriage rules in your tribe, and you will be ostracized. The punishment for breaking [L] is less obvious: you will be stigmatized as irrational, treated as subject to weakness of will or some other mental malady.

The fact that rules can be broken and that they come with enforcement provisions is reflected in their grammatical mood. Rules are often expressed in terms like "Do this" or "Do not do that." They are imperatives. There are also rules that permit, instead of precluding or requiring, although they are of less interest than the imperative ones because mere permissions don't explain as much as prohibitions or obligations. It is because rules and norms have enforcement provisions that they are facts about a community, for it takes a community to see to it that people comply and are punished for failing to comply. The explanatory power of a rule rests on enforcement of some kind or other and, thus, on a community that recognizes the rule and ensures compliance to at least some limited extent.

To see the need for sanction if a rule is really to be explanatory, consider an example. Suppose that I always wear a tie to class. You ask why. Suppose I respond, "Well, I have a rule I impose on myself: On days that I teach, I have to wear a tie." But you will rightly complain that my rule doesn't explain why I wear a tie when I teach. At most it identifies the behavior as an action. What if I add that when I violate this self-imposed rule, I fine myself $10, by tearing up a $10 bill. Does that explain my behavior?

No. In fact, now you have another question: "Why do you impose this fine on yourself?" You want to know what forces tie wearing upon me, which must be some community-based sanction. But if I say the college has a rule that professors must wear ties to class, and the fine is $10, then my behavior is explained.

Nevertheless, keep in mind that though the rule has explained my action, it can be broken at the cost of $10. Sometimes I break it and get away unnoticed by the proctors. This imperfect enforcement does not deprive the rule of its explanatory powers with respect to days when I teach and do wear a tie. That is an important difference between explanation by rules and explanation by causal laws. The rules retain their explanatory force even though they are sometimes broken, even though violation is not always detected or punished. By contrast, a generalization violated as often as a rule is would lose its causal explanatory force even for the cases that are in accord with it. But someone else's violation, even repeated violations, doesn't deprive my compliance with a rule of meaning as an instance of rule following. That's why the rule breaker's violation doesn't affect the rule's explanatory power with respect to my behavior, even if the rule breaker doesn't get caught.

In fact, imperfect enforcement may be essential to the claim that explanations of human action cite rules, not laws. Imagine a rule that was always successfully enforced, one that could never be violated without the perpetrator's being punished. Call this rule *R*. In such a case we would have an exceptionless generalization:

> Rule *R* is always followed or the violator is punished.

We wouldn't need rule *R* to explain actions in accordance with rule *R*. We could just cite the general law, for instance, that teachers wear ties or are punished. But that is just the sort of generalization that interpretationalists deny we can ever find in human action. What is more, it is just the sort of thing that naturalists search for. A perfect record of rule enforcement could be attained only by nonhuman agencies, superhuman ones, or purely mechanical ones. A human enforcing any rule would have to do so as a rule follower. But any rule could be broken, including the rule to enforce rule *R*. Thus the record of perfect enforcement could be interrupted as well. If that is correct, rules must necessarily suffer from the possibility of imperfect enforcement. But that means that showing someone getting away with violating a rule or a norm does not deprive that rule of explanatory force for someone else. That is, violation cannot deprive rules of their explanatory force, if they have any to begin with.

Only a few of the rules and norms we follow are consciously in our minds when we are following them. There are rules that many of us could not formulate correctly even though we follow them scrupulously. Other rules are so evident that they may never have formed themselves into words in our minds. Some of them are silly, like, Don't spread peanut butter on the soles of your shoes before vacuuming the carpet; others are complex, like the rule that governs the use of "who" and "whom," a rule few can state, though many obey. Nevertheless, the rules that explain our actions must somehow be represented within us. Rules and norms would not explain our behavior if we merely acted in accordance with them by accident or through the operation of some kind of causal mechanism. If rules do explain our behavior, it must be that we act "out of" a recognition of them, though it may be nothing more than an unconscious recognition. It must be the case that we could formulate the rule or norm, given the right setting and enough time and thought. Otherwise, how could rules work to give meaning to the actions they explain? Thus, rules have to have some sort of intensional existence in our minds. Moreover, any other kind of existence would make it hard for us to "break" rules—to ignore them. A wholly nonmental existence would turn a rule that most people followed into the beginnings of an unbreakable causal generalization, ready to be refined in the direction of an exceptionless law.

Though sociologists and anthropologists help themselves to the concept of rule and norm as we have articulated, it has been largely left to philosophers to attempt to produce a more perspicuous and explicit account of what a social rule or norm is. Doing so enables interpretative social scientists to identify the conditions under which they operate, and to distinguish them from social behavior that is not norm-guided, as well as reveal how they give meaning to and explain human affairs. One such theory, developed by Christina Bicchieri, makes clear how social norms and rules depend on beliefs and desires, and in fact on the way beliefs and desires explain action. And it makes clear how rules and norms differ from regularities about behavior.

A sketch of Bicchieri's analysis begins with some statement of a rule—for example, Stop at red octagon road signs, or Professors must wear ties to classes. Call this rule *R.* What would make *R* a social rule or norm that explains behavior in accordance with it? As we noted above, it is not enough that one person impose it on himself to make it such a social norm, and imposing it on oneself certainly doesn't explain action in accordance with it. Bicchieri argues that *R* is a social norm for an individual *a,* in a population of other people, if the following conditions hold for *R* and *a*:

> First, *a* knows that *R* is a rule, and that it applies in certain situations—in this case to professors teaching classes or people coming to stop signs.
>
> Second, *a* wants to or prefers to conform to *R* in these situations, on the conditions that:
>
> *a* believes that enough of the other people conform to *R* in these situations,
>
> and
>
> *a* believes that enough of the other people expect *a* to conform to rule *R*, and in some cases the other people may prefer that *a* conform to *R* and may even sanction violations of *R* by *a*.

The two parts of the second condition reflect the difference between social norms and private rules. What makes *a* prefer conforming to the rule is *a*'s beliefs about other people's beliefs and desires. "Expectation" in this analysis is not merely predictive, it can also be "normative": *a* may believe that other people not only predict that *a* will conform to *R* but will disapprove if he does not, or even punish his nonconformity.

Bicchieri notes that a rule can be a social norm even when no one in the social group is acting on it. She also wants to allow for incomplete conformity to rules. That is in part the source of the qualification to the second condition that "enough of the other people conform and expect *a* to conform." "Enough" as opposed to "all other people" allows for the incomplete conformity that characterizes all social rules, including the examples noted above. Notice that *a*'s other beliefs and wants are relevant to whether *a* conforms to *R*, and that *a* may even be mistaken about whether *R* is a social norm, if his beliefs about others are mistaken.

Bicchieri provides a nice example that illustrates how beliefs about social norms work to explain behavior even when the agent who governs his behavior by them is wrong about whether a rule is a norm: in some social groups there are a few "prudes," a few individuals who conform to a social norm against premarital sexual relations, because they satisfy the conditions above, owing to false beliefs about most other people's expectations. In fact it may be that most people in the group actually engage in such practice and use contraception so that the consequences of rule violation do not obtain. It is characteristic of norms that norms exist and have force for the prudes owing to these people's own beliefs and desires. They mistakenly believe that enough others conform to the rule against premarital sex, and mistakenly believe that these others expect them to conform to it. Thus norms can exist even when no one is acting in conformity with them, and they can still have an important explanatory role even when this is the case.

CAN WE RECONCILE RULES AND NORMS WITH CAUSES?

It is tempting to suppose that the fact that rules and norms must somehow be represented in us, our beliefs and desires, makes for a possible line of reconciliation between causal explanation and the explanation of action as meaningful. Our recognition of the rules governing our actions is in some way at least part of the cause of the action. Thus, when social scientists search for the rules that make behavior meaningful, they are also engaged in a causal inquiry, within limits. It can't be denied that deciphering the rules governing people's behavior increases to some extent our ability to predict it. How could that be, unless our learning what the rules are that "govern" an activity reveals to us at least something about its causes?

It is certainly true that learning the rules governing, say, a ritual enables us to improve predictions of the behavior of the people acting in accordance with it. So determining the meaning of actions provides some causal knowledge. But, the interpretationalist argues, these concessions miss the point of the social scientist's interest in rules. Improvements in predictive power with respect to human actions are a relatively unimportant byproduct of our study of human behavior. Our dependence on minimal causal hypotheses reflects nothing of importance about the kind of knowledge social science aims at. Understanding the meaning of a stranger's actions provides predictive knowledge only up to the limits of our own quite weak powers of predicting one another's actions. In any case, prediction is not the aim of such understanding.

Consider the way anthropologists proceed in the attempt to understand the behavior of an utterly foreign people. They begin, of course, by trying to learn the people's language, that is, the rules governing their speech acts. To do that, they must first assume that the noises the people emit are actions, that is, that the people are following the rule expressed by [L]. Whether treated as a causal generalization or a rule, [L] is pretty vacuous. That is a serious problem for a causal reading of [L], as we have seen. A vacuous generalization that can be embraced no matter what behavior occurs has no causal explanatory power. But that is no problem for the view that treats [L] as a rule usually followed but sometimes violated. In fact, if we set out to learn the foreigners' language, we must attribute [L] to them. And the only evidence that could lead us to deny [L] as a rule for these people is to conclude that their noises do not have meaning, but nothing would make us surrender the assumption that they do. So [L]'s vacuity is thus essential to its function.

To see the function of [L], suppose that, without the aid of a translator, we begin to learn our subjects' language, but our translations of their remarks always come out as falsehoods. For example, although the sun is shining, our translation of their noises comes out as, "It's raining." Someone who has just had three helpings of a dish and cleaned his plate each time says something that gets translated as, "What a revolting dish. That tasted terrible. It must have been spoiled, and it was too spicy to eat." If most of what the foreign people say comes out as false, the fault must lie with our translation of what they say. Either that or the people are fundamentally irrational—that is, we cannot identify the meanings of their actions. Since social science commits us to treating actions as meaningful, it commits us to treating people as rational most of the time. It commits us to the truth of [L] for all people. Thus the fault must indeed lie in our translation. We have not yet learned the natives' language.

Learning their language is a precondition to learning all the other rules that make foreign people's actions meaningful. Therefore, [L]'s role as a truth we would not give up, come what may, is crucial to the method of social science. As we learn more of the language, we find ourselves able to learn more of the rules that govern behavior in the society. Finally, we reach the point where we can predict every step in the religious ritual of the leading cult among our subjects. The rules we have learned make each of the actions in the ritual meaningful. They enable us to make quite precise predictions about the order, duration, topography, location of the actions. But notice, the predictions are no better and no worse than those that natives have always been able to make about this ritual. In fact, they are exactly the same as native predictions. Our anthropological inquiry has brought us to the point of knowing the detailed folk psychology of our subjects. Beyond this point, improvement is not possible, and more important, it is not necessary. For the success of our explanations is not judged by predictive success or its improvement.

The search for causes and the search for meanings part company at this point. Our own folk psychology and that of other people reach a certain level of predictive precision and stop. A naturalistic approach, which seeks causal knowledge, cannot stop with folk psychology. It must continue to demand improvements in predictive knowledge. Such improvements are the marks of further discoveries about the causes of behavior. But the improvements can only be secured, if at all, by approaches that ignore the meaning of actions. For we know the whole meaning of the action, so there is no more about it to discover. Thus there is no further insight about meanings that could help, and causes that add nothing to our understanding of meaning are of no interest to a social science.

Rule-following action is indeed caused by beliefs and desires, in which the rules are represented. But as we have seen, there are no natural laws connecting beliefs and desires with actions, neither at the level of the greatest generality, [L], nor at the level of the narrowest precision—like traffic rules. That means that when I explain someone's stopping for a red light by appeal to her beliefs and desires, her recognition of the authority of the traffic rules the light reflects, I am committed to no causal laws about traffic rules and behavior. There is no causal law about how the beliefs that the light is red and desires not to violate the law that invariably combine to cause applications of the foot to the brake pedal. I can be certain there are no laws about events that fall under these descriptions. For I know that people run red lights all the time, even people with the right beliefs and desires.

I admit that there may be some large number of generalizations in neuroscience about how my brain works that explain the movement of my body. Neuroscience may even be able to explain how my perceptual apparatus and my central nervous system work to connect the beliefs and desires to the action on any particular occasion. But the set of such laws will differ from occasion to occasion. More important, the laws will not connect the beliefs and the desires to the actions under their descriptions as beliefs, desires, and actions. They will connect them under descriptions of electrochemical and molecular changes in the brain and movements of the limbs. And such connections cannot show the meaning of applying the brake.

Of course, there is a causal mechanism underlying human action. But knowing everything there is to know about that mechanism won't elucidate the significance of actions. That is what makes even the complex underlying causal mechanism relating meanings to actions irrelevant to their explanatory function. The reliance of meanings on causes is no grounds for a reconciliation between them. All the causal mechanism does is to justify a singular causal judgment, that the desires and beliefs the agent had on that occasion were part of the cause for the action she then undertook. The causal mechanism does nothing to throw light on the meaning of what she did.

The philosophers who advanced the logical connection argument we examined in Chapter 4 provided a similar argument for the claim that causal information plays no role in explaining action, even when we accept that actions are caused. They held that to identify something as an action in and by itself precludes the very possibility of a causal explanation of it. That isn't quite the radical thesis of free will, that our actions are uncaused. It's the claim that they are not causally explainable.

The argument that causal explanation of action is impossible proceeds by example. One favored example is the action of signing a check. Every time I do that, the action is caused. But there can be no law or laws that cover every case of check signing and connect it to some invariable prior event. Why not? Because I can sign a check in a vast number of different ways: with my left hand, my right hand, my left foot, in block letters, in script, with a pentagraph, ink, pencil, red pencil, large signature, small signature, and every size in between. I can even sign electronically by checking a box on a computer screen. Any way I do it, it will still count as a check signing. But each of the different ways is a different movement of my body. As such, it will be caused by a different set of prior events. Recall the maxim of common sense and science: "Same cause, same effect." Hume, of course, made this idea the cornerstone of his analysis of causation. One event cannot be the cause of another if repetitions of it are not followed by other cases of the effect. The contrapositive of this principle operates here: different effects, different causes. But if each act of check signing is caused by a different set of prior events, the check signings must be linked to the differing causes by different laws. If there are an indefinite number of different ways of signing a check, there must be an indefinitely large number of different causes, one for each of these different effects. There must, therefore, be an equally large number of different laws required to connect them. But that means that there can be no finite set of generalizations connecting the general class of events of check signing to any other general class of prior events, no causal laws of check signing, and no causal explanations of it. A fortiori, the explanations of what I am doing when I sign a check are not causal and don't hinge on any causal regularities.

Like the logical connection argument examined in Chapter 4, this argument is too strong, but it still makes an important point. At most it shows that actions cannot be causally explained as actions. That is, once we have described a movement of the body as an action, we cannot, on this argument at least, give it a causal explanation. For the causes of what happens in and to the body do not explain its movements as an action. Every event has more than one possible correct description, and any event described as an action can also be described in terms that do not bring it under the label action. An action can be described in terms that treat it as a purely physical event. As such, it is open to a causal explanation and perhaps even a simple one at that. The last time I signed a check, the action was identical to a particular movement of my arm under specified circumstances. The movement of my arm was subject to a purely causal explanation. The next time I sign a check, the movement of my arm will differ, and the resulting shape of the signature will be different. Since the events to be explained differ, the causal explanations of these two events will differ too, in initial conditions—the prior state of my body—and perhaps also in the particular physiological laws involved. And if the number of ways the action of check signing can be accomplished is finite, then there is in principle a single causal explanation for all cases of check signing: just put together in a monster general law all the different causal laws governing each of the finite number of movements that exhaust the ways I can sign a check. To block this possibility, the number of different ways it is possible to sign a check must be infinite, not just an indefinitely large number. But the number of ways of singing a check isn't infinite, is it? So, a causal explanation for the general class of events of check signing is—in principle—possible.

But that is a hollow victory for the cause of reconciling causes and meanings. The number of ways of discharging any action is at least vast, though perhaps not infinite. Therefore, the complete causal explanation of an action will be too long and complicated to be of any use—it will not help explain an action because it will be too complicated to absorb. More important, it will not improve our predictive powers beyond those of folk psychology because it will be too difficult to establish the neurophysiological initial conditions; it will be too difficult to complete the complex calculations from these conditions to the predicted movement of the body before it actually takes place. Thus, the causal explanation of action is not, after all, logically, conceptually, and philosophically impossible; it's just physically, technologically, and practically impossible.

Nevertheless, we do actually succeed in explaining actions. And we do so by citing the rules they fall under. These rules explain because they render the actions intelligible to us. But how and why does knowing the rule under which an action falls make it intelligible to us? And what does intelligibility consist in anyway? These are two abstract philosophical questions that social scientists think they can neglect. Before the end of this book, we shall return to them and show that they cannot be avoided. Answering them brings us face to face with some central problems of epistemology.

THE SOCIAL CONSTRUCTION OF SOCIETY

Social science is more than the study of individual actions. In fact, to the extent that actions are the result of following rules, they must involve other people. For rules come with enforcement clauses, and enforcement requires others. Many sociologists and some philosophers hold that by putting together rules we learn about in examining a society, we can identify the particular roles and institutions that characterize it. More important, once we recognize that institutions and social roles are constructed out of rules, our whole attitude toward society must change, as we will discuss. Some go on to argue that, with this change, the goals of social science must change as well.

Societies differ in their institutions. But an institution is not a building or a physical entity of any sort. It is constructed and maintained by the actions of individuals. One of the things the anthropologist uncovers about the societies studied is their distinctive institutions. The anthropologist does so by identifying the rules individuals obey and determining how the rules combine with one another to generate roles, like judge, or bridegroom, or stepmother, and how the roles combine with others to generate institutions like law courts, wedding ceremonies, and adoption. Just as some theories are devoted to explaining the meaning of individual actions, others are devoted to explaining the cultural meaning of institutions. A ceremony, for example, may be illuminated as a ritual whose meaning is unknown to its actual participants. Thus the meaning of the actions of participants in the ritual may not explain it. Rather, the anthropologist must uncover the real meanings that the ritual symbolizes. These meanings may not be directly represented in the head of any participant. Thus the illumination the real meanings provide can hardly be causal explanation.

Some of the symbolic meaning of our own rituals or other institutions may strike us as obvious and uninteresting. Indeed, much of what the sociologist or anthropologist can tell us about our own society seems banal. Similarly, much of what is exciting to us in reports about another culture would seem quite banal to members of the society whose character is reported to us. For this reason, it is sometimes suggested that when we have found the meanings interpretationalists seek, we have learned only obvious platitudes and nothing really new and deep either about ourselves or about others. Interpretation, it is said, uncovers only banalities. But in identifying the roles and institutions characteristic of our society and the different ones characteristic of others, sociologists and anthropologists may have further aims beyond mere redescription of the obvious that explains it. And in these aims lies the defense against the charge of banality, as we shall see below in this chapter and even further in Chapter 8. In brief, however, showing how rules in people's heads constitute human affairs and the institutions through which they proceed, the anthropologist and sociologist reveal the conventional "constructed" nature of society and thereby show that it can be changed, revised, reformed, or overthrown. This is a far from banal outcome.

Although social scientists who argue for the constructed character of society have not been as explicit about the matter, some philosophers have tried to identify the specific ways in which all social institutions are dependent on intentionality, as it figures in language and thought. The version of this analysis that combines care and influence is due to John Searle.

Unlike natural phenomena, social institutions like money, private property, marriage, and political elections exist only owing to the beliefs and desires of human agents. The explanation for this fact about social institutions is that their existence consists in our having thoughts about them and acting upon these thoughts. Searle calls the first fact about institutions their "observer-relativity." Paper currency is such an "institution": pieces of paper would not be dollar bills and pound notes unless there were sentient agents who observed them and treated them as such, and had thoughts about them that lead to actions of various sorts. Second institutions like money or marriage have "ontological subjectivity"—"ontological" meaning roughly how something exists. Something is ontologically subjective only when its existence consists in the beliefs and desires of sentient creatures. So money and marriage exist only because of people's beliefs about and desires regarding them. Searle and other social constructivists emphasize that we can have perfectly objective knowledge about the nature and behavior of ontologically subjective things like money, marriage, corporations, and sports teams.

The question that theories of the social construction of society answer, according to Searle, is how such subjective things can exist and have influence in a world consisting only of physical things. This question is not so urgent for other social constructivists, since they don't share Searle's assumption that the only things there can be are physical things, so that humans and their minds are material objects and have no nonphysical properties. Searle means his analysis of the social construction of society to be at least compatible with naturalism about human action and the institutions it results in. But the way in which he shows how human language, reflecting beliefs and desires, produces human culture, can be accepted even by philosophers and social scientists not troubled by his insistence that all these social facts are also natural biological ones.

Recall from Chapter 4 that an intentional state is a belief or a desire or some cognate mental state that has content, is about something, is directed at an object of thought. When the intentional states of two or more individuals are directed at the same things, have some of the same content, and are also about one another's intentional states, the resulting collective intentionality constitutes a social fact, according to Searle. Collective intentionality coordinates many people's beliefs and desires, and so results in combinations of individual actions that constitute institutions. The rules governing red octagonal stop signs are a simple example: there is nothing particularly "Stop!"-ish about such signs; yellow downward pointing triangles could have chosen to express the command "Stop!" Coordinated beliefs and desires turn into an institution when they confer a function, purpose, role, on some physical thing or on some behavior of people. A thing or an action comes to have a certain *status*, the status of having a certain function for people. There is nothing more to a "status function" than people collectively according the thing or the behavior the "status" described by its function. Take an action like bending from the waist. This is a bow, and not a spasm, when and only when that status of being a bow is conferred on the movement by the beliefs and desires—the intentional states—of more than one person. The simple institution of the bow exists owing to the collective assignment to a bodily movement of a function—showing respect.

Note that functions can be assigned by collective intentionality to people, things, words, movements of the body, and many other things, regardless of their physical structure or composition. Collective intentionality even creates money without any currency or physical tokens at all. All that is required is an accounting system toward which enough people adopt the same collective intentional attitudes.

What is more, according to Searle all human institutions are built up out of nested status functions. Consider marriage as an institution: like many others, it is built up out of more and more specialized speech acts, each of which has a unique status function. First we need to collectively agree on conferring English or other languages' word meanings on certain noises that come out of our mouths. Then we need to collectively confer the status or function of *promising* on some of those words. Then we collectively confer on some promises the function of making a *binding contract.* And eventually, we confer on some contracts the function of being *marriage contracts*, and then we confer on some person and some words the status functions of solemnizing marriages by pronouncements such as, "I now pronounce you man and wife." Every institution in every culture is a (set of) status function(s) that things or behaviors have imposed upon them by collective intentionality in the form of rules or norms that enough people agree on as constituting the institution.

Searle in fact argues that all the institutions of human civilization thus exist owing to linguistic meaning conferred on noises and inscriptions by collective intentionality. At the outset of every social institution there is one or more verbal declaration made by someone accepted by enough others as having the right or the authority or the power to make the declaration. Notice how central this makes learning a language to identifying and understanding the ways in which a culture's institutions work. Whether Searle is right in this more radical view, his step-by-step construction of institutions from actions constituted by statements—what Searle calls "speech acts"—such as "I promise," for example—gives us a concrete target for any debate about the extent to which every human institution is contracted. What is more, it also becomes clear that treating institutions as constructions along the lines Searle has identified makes the prediction of exactly how they regulate or influence our actions no easier than predicting people's speech acts.

What is the source of the obligations to obey that come with these declarations, through which Searle says all institutions are created and maintained, and get their imperative force? What gives the speech acts and the social facts they create their "deontic powers" ("deontic" meaning having to do with duties and rights), as Searle calls them?

Saying "I promise," or "We the jury, find the defendant innocent," or "You're out!" is a speech act—that is, when said by the appropriate person in the appropriate circumstances, it is an action that achieves its outcome merely by being committed: all you have to do to incur the obligation to keep a promise is to sincerely say the words. To promise is to successfully commit oneself to discharge an obligation. Similarly, to find a defendant innocent in Anglo-Saxon jurisprudence is to be duly authorized to say those words, "We find the defendant innocent." The words alone, duly said, thereby insures the person against subsequent trial on the same charge. "You're out!" uttered by an umpire thereby makes a player out, and the player is obligated to leave the field. Speech creates the deontic power, the moral, political, social authority by which institutions derive their power over us.

Here we may link Searle's account of how social institutions are constructed to Bicchieri's analysis of the nature of social norms. Institutions are sets of what Searle calls "status functions." Things and actions are accorded the function of having a certain social status—being a promise or a deed of title to a car—when enough people have beliefs about enough other people's expectations and desires, especially beliefs about deontic powers; the obligation is that everyone has to act in accordance with the status functions belonging to the thing or the action. These beliefs and desires are themselves created, according to Searle, by speech acts committed by someone everyone recognizes to have the authority to create status functions by their declarations.

Now it may become clear why norms and institutions are often said to be constructions. One of the aims of social science is to probe the inevitability or contingency of social arrangements. Do social institutions represent the operation of forces beyond individual control, whose origins are not in human choice, whose continued existence is also beyond our control, and whose character controls and constrains our social behavior? Or are the features that characterize a society dependent on the choices and actions of its members, so that by choosing differently we may change society? People who hold the latter view often describe social roles that appear to constrain people as "constructions" or the results of "negotiation." The conception of institutions and roles as resulting from the interconnection of human actions governed by rules and norms is often thought to substantiate this view. Action is "voluntary." That is, we can violate the rules that give it meaning. Indeed, people acting together can change the rules that give their actions meaning. Thus, they can change the institutions that result from their interactions. New authorities can gain recognition and make new declarations, creating new status functions with new deontic powers. The task of interpretational social science is to identify what social groups and individuals have constructed.

In a clear sense then, social institutions are not "inevitable"; they are constructions, albeit often unnoticed ones, which we mistake for facts fixed and independent of human decisions.

Many social scientists believe that answering this kind of question about the nature of social phenomena is central to the aims of their disciplines. For them questions of predictive success are marginal or even irrelevant to the tasks of social science. The inquiries of interpretationalist social scientists are distinctly nonexperimental. Their focus on language is just the sort of inquiry forced upon us if we take seriously the demand that social science explain the meaning and significance of action. Of course, there are many who are likely to be impatient about such absorption with language. The results of an examination of ordinary language must perforce appear banal: telling us what we already know about our own language and nothing new. And the study of the jargon of sociology or anthropology will seem a species of self-indulgence. Both may be viewed as a distraction from the serious problems of social science. Skeptics about this approach are also likely to complain that questions about whether social concepts are constructed, negotiated, or neither are no more answerable than many of the traditional problems of philosophy. But for interpretationalist social scientists, the philosophy of language is as important as the study of differential equations is for physicists.

To say that social institutions are "constructed" means roughly that they do not exist independent of people's actions, beliefs, and desires—their reasons for acting. On one interpretation, this claim may not be controversial, for all will grant that without people there is no society, and thus no social roles to be filled by people. The claim becomes controversial when we add in the idea that people can do otherwise than what they in fact have done hitherto. They can violate the rules that constrain their actions, and they can construct new rules. That makes social institutions we may have thought were natural and unavoidable look artificial and revisable. If that is a fact about society that human agents haven't noticed, then it is far from a banal fact. Bringing it to the attention of people may have profound effects on them and society. Much of this realization follows, not from empirical investigation, but from reflection on the nature of human action, the concept of rules, and the analysis of social facts.

One of the main targets of this "constructivist" approach to social science has been naturalistic social science, which it opposes. The institutions of natural science are expressed in norms and rules demanding controlled experiment, observation, causal hypotheses, "objectivity," and predictive success as a test of knowledge. These rules have traditionally been viewed as reflections of the way the world is, not our choices about how to study it. They are the means dictated by the nature of phenomena, which determines how we can learn about the world. The constructivist holds that in fact these rules do not reflect any independent truths about nature. They are simply social constructions, consciously or unconsciously contrived and inculcated, but lacking in any foundation independent of human thought and action. We need to recognize that the scientific method, like other social institutions, is a human construction with no special claims to objectivity. Once we have recognized that, we will be freed from the mistaken belief that the scientific method is the only way to acquire knowledge or even the appropriate way to do so in social science.

Thus, the reply to the charge of banality combines two claims: First, the ultimate aim of social science is not simply to discover the rules that give the meaning of our actions but to show what the nature of social phenomena really is. Second, discovering the nature of social phenomena helps explain why naturalistic social science is so sterile. The discovery shows, on the one hand, why naturalism has failed to provide a "science" of human behavior and, on the other hand, why such a science would not satisfy our human interests in the meaning of our actions anyway.

Cross-cultural anthropology has provided a quite effective way of indicating the arbitrariness or artificiality of features of our society that people take to be immutable and unchangeable realities. Ethnologists often can show that an institution characteristic of our society has very different features in an otherwise similar culture or does not even exist there at all. By doing so, the anthropologist offers a powerful argument for the institution's dependence on our choices and actions. It should be no surprise that ethical, epistemological, and theological relativism is the chief legacy of twentieth-century cultural anthropology. Merely by showing that other societies have survived with institutions different from our own, anthropologists have called into question the "objectivity" of institutions. Now that it has become more difficult to undertake anthropological studies in Third World countries, cultural anthropologists have turned their attention to subcultures in their own societies, especially ethnic minorities. They hope to explain differences in actions, norms, and institutions in the subcultures as the result of differences in the way members of the subcultures construct their social realities.

Sometimes, however, anthropologists discover similarities in human action in different societies, similarities that, as noted above, seem best explained by assigning them meaning or significance that the agents do not themselves recognize. In the next chapter we examine the claim of some theorists to have discovered that behind the meanings people identify as explaining their actions, there is a deeper set of meanings that explains the actions in a more profound way. The existence of the deeper meanings is a powerful reply to the charge that focusing on meanings can provide only banalities we already know or don't need to know.

Introduction to the Literature

The notion that human action is explained by meanings goes back along a continuing thread of intellectual history to Plato. See his dialogue *Phaedo.* This view has animated much of the philosophy of history from the nineteenth century to the present. For a discussion, see R. G. Collingwood, *The Idea of History.* Several readings in Martin and McIntyre defend the autonomy and centrality of interpretation, including a selection from Collingwood, "Human Nature and Human History," Charles Taylor, "Interpretation and the Sciences of Man," Clifford Geertz, "Thick Description: Towards an Interpretative Theory of Culture," and D. Follesdal, "The Status of Rationality Assumptions in Interpretation and in the Explanation of Action." Critical papers in Martin and McIntyre include Michael Martin, "Taylor on Interpretation and the Sciences of Man," which seeks to reconcile a causal and interpretative approach.

Among social scientists, the autonomy and centrality of interpretation was advanced by Alfred Schutz, whose key papers are reprinted in M. Natanson's and D. Braybrooke's anthologies. Schutz's view of the primacy of ordinary concepts in the explanation of human action and their differences from explanatory concepts and strategies in natural science has animated a long series of methodologies, especially in sociology and social psychology. Some of these approaches are outlined in N. Smelser and S. Warner, *Sociological Theory.* Recent influential social science works in this tradition are M. Sahlins, *Culture and Practical Reason;* R. A. Schweder and R. A. Levine, eds., *Culture Theory: Essays on Mind, Self and Emotions;* H. Garfinkel, *Studies in Ethnomethodology;* and C. Geertz, *The Interpretation of Cultures.* D. W. Fiske and R. A. Schweder, *Metatheory in Social Science*, includes several essays by leading figures in social psychology defending this approach against methodological criticisms. R. Harré and P. Secord, *The Explanation of Social Behavior*, connects the substantive theory of interpretative social science with a philosophical foundation. Geertz's account of how interpretation provides cultural understanding is also to be found in Steel and Guala, "Thick Description: Towards an Interpretative Theory of Culture."

Among philosophers who have elaborated this idea and have made the notion of a rule central, the best-known work is P. Winch, *The Idea of a So-*

cial Science. Winch attributes this view to L. Wittgenstein, *Philosophical Investigations*. Students interested in Wittgenstein's treatment of rules and meaning are urged to start with S. Kripke, *On Rules and Private Language*, an accessible introduction to Wittgenstein's aphoristic text. D. Braybrooke's *Philosophy of Social Science*, Chapter 3, contains a condensed treatment of recent thought about rules in the explanation of action.

In addition to the works by social scientists cited above, P. Berger and T. Luchman, *The Social Construction of Reality*, is a well-known development of the notion that social phenomena consist of our collective interpretations of the actions of ourselves and others. Of special interest to philosophers of science is the attempt to apply this theory to conclude that scientific findings are social constructions themselves. See B. Latour and S. Woolgar, *Laboratory Life: The Social Construction of Scientific Facts*, and K. Knorr-Certina, *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Work illustrating the combination of constructivism and genderized Marxian interpretations of deep meaning with an attack on empiricism and naturalism in the study of science is Sandra Harding, *The Science Question in Feminism*.

Two classical works of analytic philosophy giving priority to introspective knowledge of why we act as we do are R. Taylor, *Action and Purpose*, and S. Hampshire, *Thought and Action*. Martin Hollis, *The Philosophy of Social Science* explores the prospects for reconciling interpretation and behaviorism under the distinction between explanation and understanding. Hollis also treats the once lively debate about whether there are alternative incompatible conceptual schemes, which cultural relativism requires.

Searle's fullest exposition of a philosophy of social science reflecting the ideas sketched above is *Making the Social World*. A summary is available in Steel and Guala, "What Is an Institution." Bicchieri's book-length exposition of the nature of norms is *The Grammar of Society*. Excerpts appear in Steel and Guala, "The Rules We Live By."

CHAPTER EIGHT

Philosophical Anthropology

Among social scientists, interpretative theories developed much more deeply in Europe over the past two hundred years, and in recent years have to some extent harnessed the details of interpretation and construction sketched in the last chapter. In this chapter we review the major theoretical movements in Europe that have influenced social scientists both there and in the English-speaking world.

PHILOSOPHICAL ANTHROPOLOGY

Explaining individual action through interpretation of its meaning is only the beginning of tasks of a human science, according to a European intellectual tradition that combines philosophy, history, psychodynamic theories, and other cultural influences. Some of the thinkers working in this tradition have named the subject philosophical anthropology, to indicate that it combines traditional concerns and methods of philosophy—especially continental philosophy—and substantive theories about human culture, society, economy, into a broad framework for understanding human nature. The agenda of this tradition is quite different either from the philosophy of social science as it is pursued in the English-speaking tradition, and from the self-proclaimed empirical orientation of English-speaking social sciences. Its influence beyond the European continent ebbs and flows during periods of optimism and pessimism about the prospects of behavioral and social science that is quantitative, experimental, and influenced by natural science.

A continual thread through this tradition is the standing it gives to social, cultural, and economic factors independent of the beliefs and desires of individual agents, and the strength it accords these forces in providing meaning to their actions that people do not themselves recognize. This real meaning of human events, undetected by their participants, directs individual human lives toward some outcomes instead of others.

HEGEL'S PHILOSOPHY OF HISTORY

Both the centrality of meaning to the human sciences and the notion that meaning may not be accessible to the members of a society have their origins in nineteenth-century philosophy. In fact, just as logical positivism spawned naturalism in the philosophy of social science, so the tradition of interpretationalism stems from the nineteenth-century German philosopher G. W. F. Hegel.

Hegel's approach to human action was part of a complicated philosophical system to which no introductory work can do justice. The sketch given here can indicate only the Hegelian sources of interpretation in the human sciences. At the time Hegel wrote, the sole recognized social discipline was history. Hegel's philosophy required that the study of history uncover the role of reason in human affairs. Reason moves through history, reflecting a plan for reaching a goal or purpose. Thus explanations of historical processes that reveal how reason pursues its plan are teleological in the sense of Chapter 5. Since reason cloaks its "planned" operations, individual agents do not recognize the goal that gives human history its deep meaning. Reason operates through the rational calculations and actions of individuals. But their narrow and shortsighted goals are not to be confused with the real meaning of history. Rather, deep reason works through the surface, transient meanings that people are aware of in their own actions to attain its real goals. The deep reasons give the real explanations of what people do. According to Hegel, this real meaning is the ever-increasing realization of human freedom, not in the sense of absence of constraints, but in the sense of ability to exercise human potentials fully. In his work *The Philosophy of History*, Hegel traced this ever-increasing expansion of freedom from the earliest times to his own.

Hegel's image of history passing through stages that expand the realm of freedom, from the Oriental to the feudal to the modern, was given new currency in the late 1980s. For a brief time after the fall of the Soviet Union and its collectivist client states, "the end of history" was proclaimed by some writers. This metaphorical label for the events of the time reflected the idea that freedom had finally, as Hegel had foreseen, fully actualized itself, at least to the extent that democratic values had become universal. Since that time, cooler heads have prevailed, and history is no longer deemed to have ended.

In Hegel's philosophy of history, the meaning of action to the individual is supplanted by the deeper meaning of events for the whole culture and the character of a nation. But if historical processes are meaningful, then there must be a subject to whom they have significance. So Hegel supposed that the agency to whose purposes history responds is "spirit" or "mind" or some superhuman, but nevertheless intentional, agent. This agent is not to be understood as an additional person working alongside human persons or even manipulating them like puppets to attain its ends. Rather, "spirit" is *immanent* to the actions of human agents, linking them together unconsciously to fulfill its purposes and to give the actions of individuals their deeper real meanings.

There is little of Hegel's metaphysical worldview still left in contemporary philosophies of history and social science. Indeed, during the heyday of logical positivism, Hegel's philosophy of history was identified as an example of how not to pursue the human sciences or their philosophy. Hegel's theory was deemed untestable, the notions of spirit, reason, and freedom were derided as unscientific, and the whole idea of seeking a purpose or direction in the sweep of human history was subject to ridicule. Instead, the philosophy of history focused on the narrower questions, such as whether intentional explanations of individual action are causal. In English-speaking analytical philosophy, Hegel never really recovered from this ridicule. But as logical positivism's grip on philosophy was weakened, Hegel's agenda and the approaches to the human sciences inspired by it came again to be taken more seriously.

Besides philosophers influenced by Hegel, many social scientists take seriously the notion that large-scale social institutions, events, and processes have a deep meaning, which is undetected by their participants but which the social scientist or philosopher can divine. Their approach, like Hegel's, identifies a hidden goal or purpose served by human affairs. But the goal is quite different from Hegel's freedom, and Hegel's agency—the world spirit or mind—is replaced with a less universal one. The European tradition, often Marxian, has identified social or economic class as the agency whose hidden purposes human actions serve. Recently, ethnicity, race, sexual orientation, and gender have been added to and/or substituted for class in these analyses.

The sections that follow discuss some of the theories that widen the arena of interpretation from the confines of [L] and other rules, norms, "structures" that give action its meaning. These theories hypothesize large-scale social forces, of which individual people are but unwitting instruments in the pursuit of ends that they do not even recognize. (In Chapter 9 we examine the tenability of the claim that there are such large-scale forces whose existence transcends that of the individuals who compose them.) Any attempt to seek deep meanings in social processes beyond the conscious desires and beliefs of individual agents must give an account of how such meanings are possible. Theories of deep meaning must sever the connection between meaning and the intentional states of individual minds that define what meaning is, and they need to provide a whole new account of it. Otherwise, this sort of interpretationalism must identify some agency or other above and beyond individual human beings that has the desires and the beliefs that accord meaning to their effects. Theologians will have few qualms about identifying God as the agency whose purposes give a wider meaning to human affairs. They employ it to provide a Christian philosophy of history, an eschatology in which human history is aimed at the final day of judgment. But few social scientists will go along with this sort of meaning, especially when they identify the interests of a class, race, sexual orientation, or gender as giving the underlying meaning of human events.

Opponents of such global *holistic* (from *holos*, Greek for "whole") interpretationalism demand that it provide a noncircular account of the meaning of human history. Such a noncircular account can be subject to test by prediction. They also insist on a causal mechanism linking the goals and purposes to the social processes that the goals and purposes are supposed to explain. Interpretationalists reject these demands as begging the question against their enterprise. As a result, debate between the exponents of deep meanings and the proponents of a naturalistic social science is difficult to carry on.

FREUD AND THE ANALYSIS OF DEEP MEANINGS

The twentieth century's two favorite theories of hidden meanings are Marxism and Freudian psychoanalysis. Interpretationalists find it problematic that both were probably intended by their originators to be causal theories. Each was certainly supposed to provide a predictive science of laws. Freud viewed his theory as a temporary stand-in for neuroscience. Marx identified his theory as "scientific socialism." He thought he had uncovered the iron laws of historical change. Both Freud and Marx have exercised a special fascination for those who seek meaning in human affairs, for each one developed a special theory of hidden or deeper meanings that drive human affairs but are not immediately accessible to human consciousness. Moreover, both theories embody prescriptions about what people *should* do as well as descriptions about what in fact they do. Both theorists held that once (some) people have learned the truths the theories convey, their actions will in fact change in a direction prescribed by their theories. Both of these theories also face a potential difficulty, for those who embrace them as noncausal theories of interpretation must show that Marx and Freud mistook the very essence of their own creations. Interpretationalists who follow Marx and Freud have to provide both an alternative interpretation of the theories as well as a diagnosis of their originators' errors in mistaking their theories for proto-natural scientific ones.

The literature on what Freud and Marx really meant and on the details of their theories is vast. No one should consider the discussion of these two figures that follows to be anything more than a very rough sketch, and a controversial one at that. Bear in mind that our interest is in how such theories are employed to identify the deeper meanings of human affairs. Anyone with a serious interest in the theories must look elsewhere for a fuller view (some sources are mentioned in the "Introduction to the Literature" at the end of the chapter).

Sigmund Freud's original theory was an attempt to account for individual behavior by appeal to forces he believed to be biological. Pending advances in neuroscience, Freud believed that these forces had to be described in an intentional way. He identified three components or forces driving human behavior—the id, the ego, and the superego—which determine elemental human goals and means. The id, the source of impulses, drives, and wants, is channeled by the ego, the source of information about the world and about available means for satisfying the id. The superego, which embodies the norms and rules of the agent's social environment, sets constraints on how the ego goes about satisfying the id's demands. The underlying model is clear: the id, ego, and superego work together to produce behavior in the same way that desires and beliefs work together to produce action. As a substitute for a generalization like [L], Freud introduced a "pleasure principle" and a "reality principle." The pleasure principle claims that acts are undertaken to derive immediate pleasure or reduce dissatisfaction, and the reality principle claims that in healthy individuals the immediate drive for pleasure is reconciled with the prospects of greater pleasure postponed. So far, then, only the labels are different.

However, according to Freud, much of the interplay between ego and id is unconscious, and most of the id's "desires" are unconscious as well. Behind our consciously expressed desires and beliefs, there are unconscious ones, which give deeper or real meaning to our actions. These desires are largely sexual, and the beliefs they work with to produce actions are about how the sexual desires are to be controlled and when they can be expressed and fulfilled. Psychopathology results when the interaction of the id and the ego becomes unbalanced. Neurosis reflects the unhealthy repression of the id by the ego; psychosis, the overwhelming of the ego by the id.

For purposes of social science, neurosis appears the more important notion. In fact, Freud attempted to explain the character of many social institutions and of historically important events in terms of repression and consequent neurosis. He suggested that social life could be ameliorated if we only recognized this unhealthy repression. Unhealthy repression of the id by the ego forces it to seek alternative means to satisfy its desires. The result is neurotic behavior—inappropriate and ultimately unsatisfying actions that produce unhappiness and worse. Society is in part responsible for individual neurosis because its norms and directives are impressed on the ego by the superego. In turn, social institutions can be identified as neurotic to the extent that they are the aggregated or summed-up products of the neurotic behavior of the members of society.

Freud's theory includes a prescription for treatment: psychotherapy. The aim of psychotherapy is to restore the balance between id and ego. Since the interaction between them is unconscious, therapeutic methods must focus on the manifestations of the unconscious in behavior—reports of dreams, free association, nondirective personality testing. And the therapist must interpret this behavior. The therapist seeks deep meanings by identifying the content of unconscious desires and beliefs. The cure hinges on the patient's coming to recognize these unconscious desires and beliefs. Once patients learn the deep meaning of their behavior, they will either come to accept it without distress or, more likely, surrender some of the desires—now finally conscious—as unhealthy or some of the beliefs as false. Mental health, according to Freud, requires that the id's drives—our basic unconscious desires—be "sublimated," channeled into healthy directions. Neurosis is the result of "repression"—the unhealthy imposition of false and ungrounded beliefs and unnatural social values, taboos, and restrictions represented in the individual's superego. Once the patient recognizes these for what they are, they lose their force, and behavior should cease to be neurotic.

It is clear why Freud's theory has held so tight a grip on the popular imagination for more than a century. First, it is easy to understand, for its explanatory strategy is that of folk psychology, a theory we are already comfortable with. Unlike other theories, such as Skinner's, it does not deny that human action is meaningful. Rather, it seeks a deeper meaning. Second, it is a debunking theory, one that allows us the opportunity to diagnose as unhealthy the orthodoxies of our culture. Third, it is sexy.

In part because of its popularity, Freud's theory has been the focus of controversies in the philosophy of social science over several generations. The classical objection to the theory is that it is unfalsifiable and therefore scientifically empty. Often this charge is said to hinge on the theory's therapeutic prescription. A patient is cured once he has learned the real meaning of his behavior, that is, he is cured just by learning the behavior's real meaning. Therefore, if a patient has undergone a course of psychotherapy and has not been cured, it is always open to the defender of the theory to excuse its failure: the patient has not yet learned the real meanings of his acts, the therapy is incomplete, or the therapist is incompetent. Nothing will disconfirm the claims made about id, ego, and superego. Since this attitude toward a patient is easy to adopt, in the hands of many Freudians the theory is untestable. This untestability of the essential components of the theory is alleged to deprive it of any scientific respectability.

This sort of attack is already familiar from its appearance in the discussion of [L] and other intentional theories of behavior. It has called forth two sorts of replies. On the one hand, people who wish to defend the credentials of psychoanalysis as a scientific theory (even those who hold it is a false one) point out that no theory is strictly falsifiable. Every theory is tested only in the company of auxiliary hypotheses (recall the hypotheses about the expansion of mercury in a thermometer that are assumed in any test of PV = nRT). On the other hand, philosophers and social scientists who reject the demand that their theories meet tests appropriate to natural science welcome the untestability of Freudian analysis or, at any rate, are indifferent to it. In fact, they claim the untestability shows that Freud did not really understand his own theory. For he thought it had to be verified by clinical trials.

Freud considered his theory to be ultimately reducible to neuroscience, but it is clear that such reduction is quite impossible. For one thing, the theory is evidently intentional. Though the id and the ego operate in largely unconscious ways, their states nevertheless have propositional content—represent the way the world is or could be. That is why the theory provides deep meanings and why some insist it cannot be causal. So, the criteria of appraisal drawn from natural science, like testability, are irrelevant to its assessment.

Some interpreters of Freud claim, contrary to Freud's own view, that his theory is essentially historical or historicist. To find the meaning of contemporary behavior, we must search the events of the patient's childhood. But, those interpreters note, no scientific theory is "essentially" historical in this way. Theories in physics and chemistry are completely nonhistorical. To explain all future positions of the planets, and all past ones, we need only to know their current positions and momenta, together with the laws of Newtonian mechanics. Given the laws of physics, to understand fully the motion of any particular set of physical objects, all we need is knowledge of their current position and momentum, along with the forces acting on them now.

Knowing their trajectory in the past adds nothing to this understanding. This is presumably because causes don't reach from the distant past to influence the future without going through intervening links in the causal chain. This does not seem to be the case in Freud's theory. To explain neurotic behavior, we need to know about events long ago in the patient's infancy. No amount of knowledge about the patient's current state—even the state of the patient's brain and the memories it contains—can substitute for psychoanalysis that retrieves the "real events" of childhood traumas that current neurosis results from. Moreover, unlike inanimate objects, patients can talk, and what they tell us is a source of confirmation for psychoanalytic theory that no merely physical theory can boast. All these arguments have been offered to defend the difference between Freud's theory of deep meanings and a theory in natural science. Each of these arguments raises serious questions to which philosophers of science have devoted much attention. As we shall see in the next section, many of the same things have been said on both sides of the controversy about Marxism as historicist or scientific.

If the latter-day Freudian theorist is right, then in effect Freud was profoundly wrong about the character of his theory, and he produced something quite different from what he had intended to provide. This possibility is sufficiently disconcerting that much scholarship has been invested in identifying Freud's true intentions and the meaning of his own claims about the theory. And, of course, much attention has been paid to a philosophical analysis of both his own theories and those of his psychoanalytical disciples, followers, and opponents.

MARXISM AND MEANING

The economic, political, and social theories advanced by Karl Marx had a profound impact on the twentieth century. Even after the eclipse of the political systems influenced by Marx's theories, his analysis continues to attract students seeking a deeper meaning behind social processes. Because Marx's specific predictions have been almost entirely disconfirmed, there has been strong incentive to convert his theory into an interpretive one without predictive implications. The attempt to accommodate Marx to a philosophy of social science that sacrifices prediction for meaning is even more difficult than the attempt so to treat Freud.

To begin with, Marxism styles itself a *materialistic* theory, one that identifies physical facts about the nature and prerequisites of human existence as the causes of change in beliefs, attitudes, values, roles, institutions, and whole societies. In particular, it is facts about the *modes* of production, the means people employ in order to survive and perpetuate themselves, that dictate the characteristics of all the rest of society. The means of production Marxians call the *base* or the *substructure*. All the rest—marriage rules, legal principles, moral precepts, aesthetic standards, literary styles, religious dogma, political constitutions—are parts of the *superstructure*. Thus, explanations of social institutions, roles, rules, relations among people, and so on, are to be found in facts about the means of production. Which facts are these? Ultimately, they are the ways the means of production must be organized at any given time in their development to ensure the survival of society.

Society goes through stages of social organization determined by the levels of development that the means of production have reached at any given time. For example, the shift from subsistence agriculture to modes of production that produced tradable surpluses eventually gave rise to feudalism in Europe as the dominant social organization. All the rights and duties of the feudal lord, vassal, master, journeyman, yeoman, and serf are supposed to have been causally determined by the technology of agricultural production in the Middle Ages in Europe. Elsewhere, in Mesoamerica, for example, agriculture required vast irrigation systems and highly centralized production, storage, and distribution. So, a different set of social relations emerged. In Europe, feudalism gave way to the nation-state because the feudal modes of production gave way to commercial ones.

Each of these stages is characterized by a particular form of ownership and property. These forms, expressed in legal writ and social norms, are essential to the society's survival. But like other aspects of society, they are dependent for their existence on the means of production. Property provides individuals with interests, a stake in production, and these interests clash. As the means of production shift, the classes that are endowed with property rights over these means become weaker or stronger. Eventually, when one system of production replaces another, the ruling class in that society changes. Thus, as the factory replaced agriculture, the landed classes—the nobility—lost power to the mercantile classes. In consequence, political organization changed—from monarchy to bourgeois democracy. Marx predicted that continued changes in the modes of production, economies of scale, for example, and automation, would cause further centralization in industrial organization. This centralization in the means of production would eventually shift the balance of power from the interests of the entrepreneurs to the workers, thus ushering in socialist forms of political organization.

According to many, Marx's theory, like Freud's, is prescriptive. That is, it tells us what ought to be the case as well as what is the case. It not only prescribes the transition to socialism as morally right but also holds that the acceptance of Marxian analysis by the proletarian class will precipitate the shift to socialism it prescribes. This part of the theory relies on a detailed account of the meaning of human institutions, a theory of deeper meanings than those provided by common sense. This theory of deep meanings provides the basis for ideological analysis and criticism, whose basic concept is a clearly evaluative, morally appraising one: the concept of alienation.

According to Marx, humans find meaning and value in their lives through productive activities. By creative interaction with nature, humans make their environment meaningful. Here "meaningful" does not just reflect human aims and beliefs; it also expresses approval of these hopes and purposes as valuable. But capitalism exploits labor, both in the sense that profit is unjustly derived by labor's being paid less than the value of its output, and in the sense that labor is deprived of a meaningful relation to its productive activities. Capitalism measures a person's productive work in money instead of a socially valuable product. It homogenizes relations among workers into a dollars-and-cents connection, making work obligatory instead of voluntary. It replaces the true value of work with commercialized distractions. In all these ways, capitalism makes us strangers to our own products, to our fellow humans, and to our own essence as productive agents. Making someone feel like a stranger is what "alienation" amounts to.

The means by which capitalism alienates us is the social superstructure it creates. It creates the values, ideals, laws, social norms, and institutions people live with. These features of society constitute an ideology—a "form of consciousness" that legitimizes and supports certain social institutions. Despite the fact that they do not serve the general interest, these institutions ensure the domination of society by a part of it. Ideologies often persist after the means of production have begun to undermine them; they hinder the social changes those substructural forces have set in motion; and they divert attention from these changes.

Strictly speaking, almost everything in the superstructure of society is an ideological rationalization determined by the means of production. But the term *ideology* is usually restricted to a pejorative characterization of views, both factual and normative, with which the Marxian disagrees. Once a belief has been traced to its ideological roots as an instrument for the maintenance of capitalism, the question of whether it is true—in spite of its roots—may be hard to take seriously. Once the interests served by an idea are identified, once we know that only some classes of society, say, the bourgeoisie, benefit from other classes' accepting an idea, we can be pretty confident that the idea is false, baseless, and unwarranted.

The ultimate object of exposing the ideological character of ideas, beliefs, expectations, theories, religions, and so on, is to secure a revolutionary set of social arrangements that will put an end to human alienation as well as more pedestrian forms of misery. But the immediate aim of the critique of ideology is reflexive. By revealing to ourselves both the artificiality of a social institution and whose interests are served by it, we are freed from illusions about it and from its hold on us. There is in this notion a close similarity to tenets of Freudian therapy. Once the blinders are drawn from our eyes, little or nothing more need be done to create a revolutionary consciousness, to be emancipated.

The eclipse of the Soviet Union and other states organized in accordance with a purportedly Marxian pattern of government has not led to the extinction of Marxian theories. If anything, it has removed from Marxism the burden of association with a political system characterized by grave repression and social failure. As the Soviet interpretation of Marxism as a theory of class exploitation declined, Marxian theories began to appear that identified the dominant racial or ethnic group or the dominant gender instead of the economic class as the agency of political and economic exploitation. Thus, feminist theories offer an interpretation of human social relations that makes them meaningful as a struggle for sexual equality or justice. Such theories are often harnessed together with similar approaches to social processes that interpret the latter as a struggle between European colonialism and non-European racial and ethnic majorities and minorities throughout the world. It is a matter of some irony that all these strands of Marxian analysis continue to reflect the Hegelian idea that the meaning of history is to be found in the inexorable march toward the goal of complete human freedom.

Marxian theory has been at least as fruitful a source of methodological problems for the philosopher of social science as Freud's psychoanalytic theory has been. In fact, many of the same questions that daunt Freud's theory trouble Marx's and his successors' theories as well.

Both Marx and Freud advanced theories they held to be causal, yet both traded heavily in intentional notions. So both theories have been the scene of debate about whether intentional theories can be causal. Both have also been targets and benchmarks for claims about testability and falsifiability, independent of the problems that intentionality makes for such requirements on scientific theory. Just as Freud's theory is accused of being held true come what may, Marx's theory is subject to the same accusation. Marxism's most important predictions have been disconfirmed. These include predictions of the "immiseration" of the proletariat in capitalist society—the economic position of the proletariat is always getting worse; about the place and circumstances of socialist revolution—in countries where capitalism is most highly developed; about the creation of socialist man—through persistent planning, indoctrination, and exhortation, as in Cuba. None of these predictive failures seem to have shaken the conviction of its proponents very much.

Marxism was for a long time the focus of much philosophical analysis. The reasons are obvious: its political importance and its claims to lay the foundations for "scientific" socialism. Even more than Freudianism, Marxism was a theory to which philosophers of science applied their tools and conceptions in order to demonstrate that it was or was not really scientific. During the heyday of logical positivism, the verdict on both theories was largely negative. For positivism requires that meaningful statements be in principle verifiable or falsifiable, and neither theory was so treated by its proponents. While Marxism's theory was stigmatized as both unverifiable and unfalsifiable by positivists, others attacked it as false or unscientific or as the result of an unscientific research program.

For example, Marx's labor theory of value, in contrast to bourgeois microeconomic theory, denies that the value of a commodity depends on the strength of market demand for it. Instead, Marx argued, the amount of labor that goes into the production of a commodity determines its value. All profit is thus the result of exploitation, for it is the difference between the wage rate and this labor value. The trouble with Marx's theory of value is that there seems to be no way to measure the value of the labor that goes into production except by its wage. Without a means of calculating the "real" labor value of commodities, it is difficult to apply or test Marxian theories about the nature of capitalistic exploitation and its future. In fact, most of the suggested ways of measuring the value of labor independent of its wage are either unsatisfactory or lead to the disconfirmation of Marxian claims about capitalistic exploitation. The continued commitment of Marxians to the labor theory of value is usually criticized as unscientific because it is impervious to evidence—favorable or unfavorable. When expressed in testable form it appears to be false in many cases.

Another issue to which philosophical reflection has been devoted is Marx's claim about the historical inevitability of successive stages of economic and social organization, and the laws that determine these stages. The alleged inevitability of socialism and the unavoidability of prior stages of economic development, such as capitalism, are likely to breed both resignation and complacency among revolutionaries. That is because their actions cannot hasten the revolution. It will occur when the time is "ripe." Moreover, the substructural determinism of Marx's theory makes it difficult to see how Marxism can even be "thought of" by members of the exploited classes, and influential among them before the breakdown of capitalist society. The theory requires that agents be free enough from the constraints of their economic substructure to embrace Marxism and to act on it. Yet if such freedom from the substructure is possible, then Marxian claims about the determination of thought and action by the means of production are jeopardized. The problem remains how to reconcile the deterministic element in Marx's theory with its prescriptive force.

As briefly noted above, both Marx's and Freud's theories are sometimes described and defended as historicist ones. A theory or method is historicist roughly if it holds that in order to understand and to predict subsequent states of a system—whether a whole society or an individual person—we must have detailed knowledge of the (usually distant) past states of the system. Even to predict the very next "stage" in the development of a neurosis or an economic system, we need to know about events long past in the life of the individual (usually the patient's infancy) or the society—sometimes even its prehistory.

Naturalists and other empiricists have strong objections to historicist theories. As we said above, the problem such theories raise relates to causation. In astronomy, all we need to predict and explain future states is a description of *present* ones, plus the dynamical laws of mechanics. Given the present position of the planets, we can predict all their future positions—and retrodict their past positions too. We don't need to know their past positions to predict their future ones because all the causal forces determining future positions are present and detectable in the current state of the system. In other areas—biology, for instance—we sometimes need to know about the past in order to project the future, but that is presumably because we do not know the dynamical laws that govern the systems under study. It's not because the distant past continues to exercise an independent causal force on future states.

When Marx's or Freud's theories are described as essentially or unavoidably historicist, what is meant is that past events really do continue to exercise control over future ones. Past events do so, not just because they have brought about the present state, but because no matter what the present state is like, future states could only have been brought about by a certain history. In certain biological sciences, such as embryology, we need to study the past states of an embryo in addition to the present state, because we don't know what features of the present state will determine the next state. According to historicists in the development of society or personality, it's not just ignorance of these present factors that makes the past important. Over and above them, there are causes in the distant past.

Thus, in Freud's theory, we cannot use the adult effects of infantile experiences to predict psychopathology. We must identify those early experiences themselves. In Marx's theory, in order to determine the future of a society, we need to know the particular stages through which it has passed. Studying the effects of past stages on the present is not enough.

This sort of causation bears the same problems as teleological causation. Recall in Chapter 5 ("Causation and Purpose") the problem of future events, events that don't yet exist and therefore cannot bring about present ones. Historicism requires that past events, which no longer exist, bring about future events somehow without affecting present ones. But if past states do not leave a mark on the present that we can identify and employ to chart the future, then their determination of the future cannot be through causal means known to the rest of science. For causation does not work through temporal gaps any more than it works through spatial gaps. There must be chains linking the earlier to the later. And a complete knowledge of the intrinsic causal properties at any link, together with laws, should be enough to determine the character of future effects, without adding information about earlier links.

If understanding and predicting the future requires unavoidable appeal to events of the distant past, then the explanations such knowledge involves will not be causal. That is another reason why philosophers of science have raised grave doubts about Freud's and Marx's theories, when interpreted in the way their originators wanted to treat them—as scientific theories.

By and large, the debate in the philosophy of science came out against the conclusion that Freud's and Marx's theories were scientific in the strictly positivist sense of that term. Therefore, positivist philosophers read them out of the corpus of legitimate knowledge. Nevertheless, the same arguments were employed by antipositivists and exponents of interpretative social science as reasons why these theories are not to be judged by the standards of causal theories. They argued that the study of deep meanings that Freud's and Marx's theories reflect need pay no heed to strictures inappropriately drawn from natural science. After all, faced with the choice between surrendering the intelligibility that Marx and Freud—and common sense—provide to human affairs, and ignoring the strident claims of logical positivists, the antipositivists and interpretationalists thought it obvious which alternative should be endorsed. Thus, a countercharge arose against positivist philosophers of social science and social scientists eager to be as "scientific" as possible. When called an advocate of pseudoscience, the defender of meanings labeled the opposing school "scientism"—an exaggerated and ideologically explainable respect for a certain mistaken image of science. Indeed, two of the most remarkable figures to have been in thrall to "scientism" were Freud and Marx themselves. Their own theories must be reinterpreted in order to free them from this incubus.

FOUCAULT AND BOURDIEU ON THE CULTURAL SUPERSTRUCTURE

Marx's notion of an ideological superstructure of social and cultural institutions, practices, norms, values, even styles and fashions and discourses—ways of describing things—has had an influence in the social sciences, especially in Europe, that lasted long after the influence of his economic theory waned. In the research programs of sociology, history, and anthropology that sprang from it, the two most important figures were Michel Foucault and Pierre Bourdieu.

Foucault was a philosopher trained in the French tradition and much of his later work in epistemology and metaphysics has no direct bearing on social science or its distinctive philosophical problems. But his earlier and more accessible works made him a household name among social scientists with a revolutionary social and political agenda in the last third of the twentieth century. Foucault brings together the notion explored in the last chapter that social and cultural facts are constructed, artificial, revisable, and not inevitable, with the Marxian notion that they have a life of their own, independent of the beliefs and desires of individual human agents. In fact, they control the beliefs, desires, and consequently the actions of individuals through processes of which the individual is not aware. In coming to grips with Foucault's theories, it is important to continually ask how something can on the one hand be constructed and revisable, and on the other hand be overpowering and undetectable in its operation.

Over the course of twenty-five years, Foucault advanced a historical analysis of changes in central concepts of social life: mental illness, imprisonment, sexuality. The meaning of these concepts changed significantly over historical epochs, especially in Europe. Uncovering these changes, Foucault held, reveals social structures, mainly ones that control individual behavior. They do so by giving acts meanings that the agents who perform these acts don't recognize. There are thus meanings that exist independent of anyone's recognizing them, and interpretations of meaning that trump the individual's own authority about the meanings of their actions.

In his adoption of the hypothesis that there are structures of meaning independent of individuals, Foucault was part of a tradition of social science that echoes Hegel's idea of reason operating in history independent of people's reasons. In the commitment to "structures" there is a strong echo of the thought of Emile Durkheim, the founder, along with Weber, of modern sociology. (For much more on structuralism and Durkheim, see Chapters 9 and 10.) The notion that such structures exert control over individuals, usually in the interests of a ruling class, goes back directly to Marx. But unlike Marx, Foucault did not explicitly ground the ideological superstructure of mental illness or prisons or sexual morality whose real meanings he claimed to uncover.

How might structures of meaning that individuals were not even aware of constrain their choices and behavior? One attractive answer to this question that finds support in Foucault is the notion that the meanings of the very words people used to think about and describe their actions impose on their thoughts boundaries that define what is possible to do, channels in which to respond to events, mutual expectations that constrain people as closely as prison walls or insane asylums. Here there are echoes of Searle's account of "speech acts" as the source of social constraint. How are such structures of meaning to be uncovered? Foucault's approach involved what he labeled "archaeology," a word that appears in the title of one of his most influential works. Largely through an analysis of canonical literary and legal texts, and other contemporary documents of the relevant historical periods, Foucault sought interpretations that revealed the real meanings of rules and norms that restricted the way people could even imagine their lives, personal relations, and roles in institutions.

He began with the notion of madness, one which from the perspective of the twenty-first century has presumably shown progress and increasing enlightenment from the notions of demonic possession in the sixteenth century, through genetic illness in the nineteenth, to social pathology in the past century. Foucault read this history as one that changed the meaning of madness as a concept people employed to marginalize and control deviance, all the while ensuring its continued role in social control. The next target of his archaeological excavations could be seen to have emerged naturally from his histories of madness: the medical clinic with its self-professed scientific basis for treatment of illness—physical as well as mental—is the result of a similar pattern of change in meanings over time, changes that are hidden from participants, though not through anyone's intentional suppression of the truth about them.

It is the social structures as constituted by these meanings that have a life and force of their own, and change in ways that maintain relations between people in tracks that ensure the society's persistence. Eventually Foucault moved on to a study of prisons, the most obvious institutions of social control. He sought to show that through all the changes and reforms, all the humanizing and all the increasing scientific approaches, and despite the self-pronounced rejection of retribution and punishment, prisons continued to be simply the most concrete manifestation of social control effected by structures of unconscious thoughts and meanings, which give the content of words, categories, and all the other ways we view social relations.

In fact, the increasingly "scientific" character of arguments offered about how society needs to be organized reflect for Foucault the inextricable interplay of power and knowledge. And this idea that social power and scientific knowledge are aspects of a single unified force has been among Foucault's most lasting influences in philosophy and the study of human affairs. Notice that this identification is not different from the epistemology of the otherwise very different empiricist and positivist approach. For these philosophers the test of knowledge, perhaps even its very essence, was its predictive power—exactly the sort of power one needs for control. The steps toward identifying it are neither very great nor entirely metaphorical.

After madness, illness, and the punishment of criminal acts, Foucault turned his philosophical scrutiny toward sex. Here Foucault faced problems similar to those that confronted Marx: how to reconcile the force and determinism unconsciously imposed on people by social structures with the individual human freedom and autonomy that enabled both Foucault, and Marx before him, to uncover these hidden substructures. Even more pressing is the problem of explaining how the discoverer of them and the rest of us can free ourselves from these constraints. It was obvious from Foucault's study of the repressive sexual norms of contemporary society that he believes we can do so. But for a thinker like Foucault the problem may be graver than for Marx. There is no internal problem within the Marxian tradition of securing the scientific tools to uncover the economic substructure that generates the ideological rationalization that most people unreflectively internalize, not knowing it is just a tool of the ruling classes' continued exploitation. But Foucault's problem is that the very tools required to undertake the archaeology of knowledge may be subject to the same analysis and critique he used them to make. The means that reveal the meanings that unconsciously constrain individuals, and may enable us eventually to overthrow the constraints, are themselves interpretative instruments to which his own analysis must apply. May we not simply always exchange one system of internalized meanings that unconsciously govern us for another, and never free ourselves from autonomous structures embedded in meanings?

Pierre Bourdieu was slightly younger than Foucault, and moved from philosophy into a more recognizably sociological tradition than Foucault, one whose influences include the great names of French anthropology and sociology from Lévi-Strauss back through Marcel Maus to Durkheim. (For more on this naturalistic structuralist tradition, see the next two chapters.) But, as we will see, he shared with the philosopher the approach to distinctive social superstructures inherited from Marx. Unlike Foucault, Bourdieu did not sever its connection to an economic substructure, but for him culture was a relatively autonomous and independent force shaping social life.

The influence of Marx on Bourdieu is evident, for he took from Marx as the central theoretical concept in the understanding of social processes the notion of *capital.* But for Bourdieu capital is not merely or even largely economic or financial or material, it is *cultural.* This sort of capital is a productive resource conveyed to individuals by their places in a society—their family circumstances and upbringing, their education, their social milieus, the symbols and signs of their professional lives, domestic lives—and the way they, in turn, start over the process that accorded them cultural capital at the outset of their own lives. Differences in cultural capital characterize individuals, and individual subcultures, professions, societies, groups, et cetera. Within such groups there will also be hierarchical differences characterized and constituted by differences in cultural capital. Marx seemed to treat ideology and culture as relatively dependent rationalizations for real differences in power, which were all, at bottom, economic.

Bourdieu accorded independent force to culture, and argued that it was the quantity and character of one's cultural capital that determined one's social power. Like Marx, Bourdieu treated classes as the significant factors in society, but he substituted culturally characterized ones for Marx's economic classes. And he provided a theory of how individuals come to internalize the cultural values and meanings characteristic of their classes, often by subscribing to aesthetic, gustatory, sartorial, and other tastes associated with the class into which they are born or raised. Cultural capital, according to Bourdieu, comes in a limited number of major forms: certification by recognized cultural institutions such as universities; a range of objects consumed by distinct cultural classes—the *New York Times* in the United States, or *Le Monde* newspaper in France, versus the English Premier League football replica kit in the United Kingdom—and, of course, the habits and dispositions of members of distinct cultural classes—going to the opera or going to the Grand Ole Opry. Bourdieu also argued that, unlike inherited wealth, for example, it is much easier to mistake cultural status for earned, deserved outcomes of individual effort and ambition, when in fact they are largely fixed and socially inherited, and unearned.

Like Foucault, Bourdieu is committed to the notion that the factors that control individual behavior are rarely ones of which people are conscious nor ones of which they are introspectively aware as they make choices and engage in social behavior. But unlike Foucault, Bourdieu held that culture shapes behavior via habits, dispositions, abilities, and other unreflective ways of interacting with others. Like Marx, and unlike Foucault, he was not committed to an approach to human affairs that emphasized interpretation of social structures in terms of meanings, hidden or otherwise, that dominated individual thoughts and actions. In common with Marx, Bourdieu saw his theory primarily as a set of causal claims about the role of cultural capital, and as such they have influenced sociologists whose approach is much more empirical than that of the followers of Freud, Marx, and Foucault.

But, like Foucault, and for that matter Marx and Freud, Bourdieu believed that recognition of the ways in which substructures or superstructures, conscious or unconscious forces, shape individual lives, the relations of classes, and the distribution of power in a society provides a basis for changing that society, improving it, liberating its members. The constraints society imposes on individuals and classes are, in the sense discussed in the last chapter, constructed, not inevitable and natural, even if they are not consciously constructed. As such, recognition of their character enables us to see that they can be reconstructed, altered, changed, broken down, or replaced, presumably by improved structures, or by no structures at all.

CRITICAL THEORY

Critical theory labels an approach to social science, its aims and methods, and the uses to which it ought to be put, developed largely by the German contemporary of Bourdieu and Foucault, Jurgen Habermas. Like these thinkers, Habermas was heavily influenced by Marx and Freud, and like them described himself as a socialist or as sympathetic to socialism. Unlike these French thinkers, Habermas was much more interested in combining the tradition of thought that stems from Hegel with English-speaking empiricist philosophy of science. Like continental thinkers, he remained committed to a visible public role for philosophical anthropology in assessing, reforming, and reconstructing social, cultural, and political life. This seemed an urgent task perhaps, in post–World War II Europe, and especially Germany. But it continues to animate critical theory and the other continental traditions in philosophy.

Like other nonempiricist philosophers, Habermas holds that there are several distinct forms of knowledge, each partial and incomplete, but all required to attain important human social ends. The most well entrenched and obvious of these is the sort of knowledge that is produced by empirical science. We understand well how such knowledge is produced, though of course its actual production is difficult, demanding, and not the result of merely applying recipes. The litmus test for such knowledge is of course prediction and control. Dangers emerge for human values when this sort of knowledge is demanded about human action, behavior, society, and culture.

Empirical enquiry in human affairs may in fact provide predictive power, but it is not the sort of knowledge about human affairs we really need.

Another equally important sort of knowledge we require is normative, morally informed, or "practical" knowledge. Habermas's and others' use of this term derives from the eighteenth-century philosopher Immanuel Kant, whose great treatise on ethics was titled *The Critique of Practical Reason*. Such knowledge is required for individual self-understanding, and to construct, value, and enhance human relationships.

Like Foucault, Habermas recognized that knowledge and power were closely linked. Each of the three forms of knowledge comes with an "interest"—an agenda for the application of the knowledge in order to attain some outcome. The inseparability of knowledge and human interests is manifest more clearly in the third distinct sort of knowledge people require: this is the understanding of how to liberate ourselves from social institutions that operate in the interests of others, how to emancipate ourselves from social constructions that are artificial, inappropriate, obsolete, alienating, and inhuman. In particular we need this kind of knowledge to free ourselves from social institutions designed from the perspective of the first sort of knowledge—that which enables one group, usually a small one with selfish interests, from controlling another group, usually larger. Here the influence on Habermas of Freudian analysis as liberation of the self from neurosis is evident. By designing institutions guided by predictive knowledge—or more often rationalized by its pretenses—bureaucracies, corporations, and government security services limit people's freedoms in ways that bureaucrats and bosses themselves do not often recognize. Usually acting out of benevolent motives, with the aim of ameliorating [improving] social conditions, individuals in authority mistakenly apply the first sort of technological, scientific knowledge to producing engineering solutions to human problems. Central to critical theory is the critique of this sort of error, and the recognition that the sort of knowledge required to deal with human problems is the second sort of moral, political, "practical" knowledge, which can only be secured through discussion, deliberation, conversation, and democratic political processes. In its place, natural knowledge and even rational planning have a useful role in social life. But, according to Habermas, limiting it to its appropriate role is a task for the third sort of knowledge as critical reflection.

The latter two forms of knowledge—practical and critical—take interpretational form according to critical theory, seeking hidden meanings among social roles, practices, and institutions, and constructing new meanings in democratic deliberations by a democratically constituted public. The outcome of this deliberation in Habermas's conception bears haunting similarities to Hegel's idea of reason, rationality, finally achieving its full expression. Here rationality is more than the means/ends efficiency and appropriateness that economists prize. Reason here takes on an evaluative, ethical significance, in which each individual acting on reason is free, autonomous, and authentic in achieving his or her own interests, ones that are shared in common with everyone else.

Along with the echo of Hegel, Habermas shares with English-speaking philosophers such as Searle a commitment to the role of language and especially speech acts as constitutive of human institutions, owing in part to the essential role communication between people plays in all human institutions. The emancipation that critical theory aims at is a "free-speech" situation in which all participants will ultimately agree, owing to the absence of deforming and controlling forces that limit human reason from its proper function.

Habermas is not the only figure to adopt the label critical theory for a philosophy of social science. But in the second half of the twentieth century it was his name that was mainly associated with it among English-speaking philosophers. There is one important fact about human affairs and individual people's relationships to the social science that other critical theorists shared with him, and that marks the movement as distinct from other traditions of philosophical anthropology and most empirical social scientists: the reflexivity of social sciences. People can come to be aware of the findings of the social scientist about them and their behavior. Inanimate objects and nonsentient creatures cannot. A science of human behavior must accommodate this possibility: it must be, to use a term from the grammar of certain verbs, "reflexive." A science proceeding by positivist methods cannot be reflexive.

As we shall see, there is a fairly well-known problem of "reflexiveness" in social science, and it is a potentially serious obstacle to predictive success in the individual disciplines. But it is not clear that the problem is either insoluble or the result of a misguided commitment to methods appropriate only in natural science. And that suggests that something more is involved in the critical theorist's claim that social science is reflexive and natural science is not.

The phenomenon of reflexiveness is easily illustrated. An economist surveys farmers' resources and opportunities and the current price of wheat; plugging this data into her theory, she predicts that there will be a surplus this fall and that the price will fall. This prediction, circulated via the news media, comes to the attention of farmers, who decide to switch to alternative crops because they now expect lower wheat prices. The results are a shortfall of wheat and high prices. Here we have an example of a suicidal prediction. The dissemination of a physical theory has no effect on its subject matter, but the dissemination of a social theory does. It is reflexive (recall the notion of reflexive verbs describing things one can do to oneself).

If the prediction had not been disseminated, it would have been confirmed. Therefore, in making such predictions, the social scientist must take into consideration a variable reflecting the degree to which the prediction will become known to its subjects and the degree to which they will act on it. Doubtless, this complication makes prediction more difficult, but not impossible. Indeed, in recent years, some economists (known as rational-expectations theorists, because they attribute rational expectations about the future to the subjects of economic theory) have held that the failure of governmental policies based on macroeconomic theories is the result of reflexiveness. Well-informed people are acquainted with the same theories the government uses. So when the government inflates the currency to make everyone feel richer and spend more, well-informed people are not fooled and the government's policy does not work. People change their behavior in ways that falsify the theories used to shape economic policy. Some rational-expectations theorists hold that reflexiveness makes macroeconomic policy (and even theory) impossible.

But critical theorists have something much different in mind when they claim theory should be reflexive. What they really seem to mean by their demand for "reflexive" theory is that social science must combine at least two of the types of knowledge that Habermas identified: descriptive, predictive, positive, and "practical," normative knowledge, which has a moral dimension. It would not merely describe the way the world is but would provide positive guidance about the way the world ought to be. Such a theory would be reflexive in a stronger sense than the one we have just discussed. Its dissemination not only could affect action, but like Freud's or Marx's, it *should* also effect action. It would prescribe the direction in which action should be taken. Not just Habermas, but most critical theorists rightly insist that methods employed in natural science could never produce such a theory. For the theories of natural science are limited to description and make no claims about prescription.

Critical theorists hold that the aim of social knowledge is not just to reveal the causes of human actions. It must also provide participants in society enlightenment as to the true, deep meanings of their actions. By doing so it would enable us to *emancipate* ourselves from false beliefs about the nature of society and the morally unacceptable effects of those beliefs on people. We shall return to the normative dimension of critical theory in Chapter 14. For the moment, we need to focus on the deeper meanings that this theory finds in human actions.

It is characteristic of many contemporary theories in social science that seek to find the meaning of human action and of human events that they embody moral dimensions, which are conspicuously missing from theories in natural science. Moral dimensions are also absent from theories about human behavior that either avoid intentional explanations, such as behaviorism, or are embarrassed by them, such as modern economic theory. This is what makes this sort of knowledge incomplete and limited when it is not absolutely exploitative of human potential. Habermas and other critical theorists held that the source of this prescriptive element in interpretational theories is to be found in the intentionality of social explanations. It arises from the language in which agents express the rules that guide, explain, and justify their behavior. This is what makes Habermas's "ideal speech situation" the goal of practical and critical social science. If critical theorists are correct, moral normativity is an inescapable dimension of folk psychology. It must explain actions by describing them in morally evaluative terms. These terms figure in the rules people act on and are the starting point of the social scientist's inquiry.

Consider [L]. It is a rule that helps determine what counts as rational. But it is not all there is to rationality, for to describe an act as rational is to praise it. What is more, rationality extends beyond the appropriateness of means to ends that [L] reflects. Rationality assesses the appropriateness of ends themselves as permissible or obligatory—or as prohibited (because they are irrational). Thus, a morally neutral social science is impossible: a science that treats people as moral agents must either wear its moral commitments openly or disguise them.

Critical theorists can identify a good example of morally contemptible deception or, better put, self-deception, in the twists and turns of the bourgeois economist's theory of rational choice. Recall that the shift (described in Chapter 6) from cardinal and interpersonal utility to ordinal utility precludes the economist's making interpersonal comparisons of welfare. Among economists, this self-denying ordinance has ossified into the claim that such comparisons are unintelligible. This belief goes hand in hand with a strict injunction, in economic methodology, of value neutrality: economists can describe, but they cannot evaluate. Their discipline is positive, not normative. Here the overt rationale is the desire to be scientific, like natural science. But the covert meaning of the shift away from interpersonal utility that economic theory effects is quite different and essentially normative. First, the surrender of interpersonal comparisons effectively blocks arguments for redistribution, expropriation, and other tactics for improving social welfare. Accepting this theoretical restriction has great social consequences and represents a moral decision that welfare inequalities are permissible. It does that by making the dubious claim that they are indeterminable. By attempting to adopt the value neutrality of natural science, economics is trying to absolve itself of the obligation to search for a replacement for cardinal utility theory, one that could make judgments of manifest injustice. This tactic enables economics to turn the normative concept of rationality into a positive, descriptive shell. Rationality is no longer able to decide between the rationality of ends. It is restricted to assessing the rationality of means, when it has any intentional content at all.

This kind of unfavorable appraisal of alternative conceptions of social science is characteristic of critical theory and of other interpretative philosophies of social science. They combine a methodological and a moral critique of empiricist philosophies. But the moral critique rests crucially on finding meanings in events that the participants did not see in them. No critical theorist accuses F. Y. Edgeworth or Vilfredo Pareto, the founders of ordinal utility theory, of consciously adopting the theory because of its prospects for rationalizing the political status quo of early twentieth-century Britain. Nor do they suggest that the developers of modern welfare economics were engaging in a nefarious game to obstruct the forces of egalitarianism. Though critical theorists seek in such results the deeper meaning of these economists' work, they do not seek the meaning in the conscious intentions of individuals. But critical theorists do expect that once people become aware of the real meaning of their actions, their actions will change. Social science is reflexive.

What kind of meanings can explain human action that are not already known to the agent or reflected in folk psychology? The meanings critical theorists need are the ones identified by Marxian and Freudian theories. So, critical theory has helped itself to much of both these approaches to human affairs. That has meant that critical theorists had to defend the intellectual credentials of theories like Marx's and Freud's.

Critical theory, despite its defense of Freud against charges of pseudoscience (by rejecting empiricism beyond the natural sciences), is not really wedded to the details of Freud's theory. Infantile sexuality, the Oedipus complex—these details are less important to critical theory than the Freudian conception that some beliefs are unfounded illusions foisted upon us by others, intentionally or not, and from which we must emancipate ourselves. Crucial also to critical theory is the technique of psychoanalysis, which the critical theorist wishes to adapt to identifying and curing social ills. Just as Freud's theory requires the analyst to identify a "text"—a story—in the apparently random and undirected behavior of the patient, so too the critical theorist examines unintended features of culture to find the real meaning of social institutions. By identifying this meaning, the critical theorist hopes to produce changes in culture the way a psychoanalyst produces changes in a patient.

The critique of ideology is probably the most important thing critical theory has taken from Marxism. It is illustrated in the critical theorist's attack (outlined above) on modern bourgeois economic theory. By means of ideological analysis and critique, we can discover our real interests and the interests of those who encourage ideological delusion even when they themselves do not realize what their interests are. And the critique of ideology can, of course, be severed from the details of Marxian theory. By substituting factors like race, sexual orientation, and gender for economically exploiting classes, the critical theorist can update Marxian theory and sever its connection to a failed political tradition.

RADICAL INTERPRETATION AND CONCEPTUAL SCHEMES

We can now understand the intellectual geography of the philosophy of the social sciences. At one extreme of the spectrum are the few philosophers and social scientists who have held that meanings can't be causes, that the knowledge social science seeks must be causal knowledge, and that therefore we must turn our backs on meanings. At the other end of the spectrum there have been a larger number of social scientists and philosophers who have agreed that meanings cannot be causes, but that they provide knowledge, so that the aim of social science cannot be causal knowledge. Between these camps has stood a body of writers advocating reasonable compromise. They hold that the arguments of neither side are right, that there is no logical incompatibility between meanings and causes. Thus, there is no forced choice between intelligibility and prediction, and we can have them both, in principle. If we have rather more intelligibility and less prediction—well, this is a practical problem of complexity and difficulties of research. It is surely not a problem that demands a complete reworking of empiricist epistemology.

Reasonable compromises are always unsatisfying. Those who demand improvements in prediction will insist that their opponents show that meanings really do provide intelligibility, not just a psychological feeling of curiosity satisfied. What after all is the test of intelligibility? Is it really just the feeling that were we in the place of the agent whose behavior is explained, with the same beliefs and desires, or in the grip of the same neurosis or ideology, we would do the same thing? Why should this feeling be a mark of knowledge? How can a feeling certify the explanation offered and the principles behind the explanation as true? Feelings are a notoriously poor guide to knowledge, as Freud or Marx would themselves attest. We feel nothing more firmly than that we are at rest. Yet this belief, despite the feeling that it must be true, is quite false—we are hurtling through the universe at a high velocity. If we had allowed the feeling of being at rest to serve as a mark of knowledge, we should still be in the grip of the Aristotelian theory of motion that the Catholic Church tried to protect from Galileo. Is it possible that the feeling of intelligibility is equally mistaken, that it can be explained away, just as the feeling we are at rest was explained away?

The real test of intelligibility, this argument continues, rather than a feeling, is the application of the information that produces the feeling to successful prediction. This application certifies the information as reliable, as knowledge. As for the feeling of intelligibility—well, it is at best and only sometimes a by-product of such successful prediction.

The fundamental assumption in this argument is epistemological: it is the claim that propositions count as additions to knowledge only if there can be independent objective evidence for them, evidence based on observations of phenomena independent of our feelings and thoughts. This is some sort of empiricism. And it leads inevitably to skepticism about an interpretative science of human action.

As we saw in Chapter 4, all the information available to us by the observation of behavior will still not enable us to choose unambiguously between two different combinations of belief and desire, combinations that have exactly the same vague and imprecise consequences for behavior. If meanings are ultimately matters of what is in our heads, then observations of human behavior alone can never decide between competing interpretative theories, whether Marxian, Freudian, commonsensical, or other, not even in principle, even when everything else is known about the rest of science (including neuroscience), and when all the data about behavior is in.

This is a claim that originated with the important American philosopher, W. V. O. Quine, and that he labeled “the indeterminacy of translation.” We are invited to imagine ourselves as anthropologists seeking to translate the language of a people with whom we share no linguistic history and no translators. All we can do is watch and listen to someone willing to speak in our presence. Suppose a white rabbit with one ear longer than the other passes in front of both of us, and the speaker whose language we seek to translate utters the noise “gavagi.” Shall we take this to mean, “There is a rabbit before me,” or “Damn it, I didn’t bring my rabbit gun,” or “White animals are rare in these parts,” or “I wonder why one of its ears is longer?” or any of an indefinite number of other candidate meanings of the utterance? Well, presumably, as we listen and watch, longer and longer, we will be able to rule out many candidate meanings. What is more, in seeking a translation of this language we will have to assume that by and large the speaker has roughly the same beliefs and many of the same “basic” desires as we do. If our translation of their words leads us to attribute beliefs or desires to some speakers that are irrational, like, "That rabbit looks so much like me that we must be twin brothers," or "I am disappointed that that rabbit did not turn into a prime number," then obviously our translation manual, or dictionary, must be faulty. This constraint on translation Quine dubbed "the principle of charity."

Now if we accept them, the doctrine of the indeterminacy of translation and the principle of charity have important consequences not only for anthropology, and interpretational social science generally, but also for fashionable doctrines about cultural relativism. If Quine is correct, then it is possible for two translators to produce two incompatible dictionaries that are each compatible with everything the speakers of the strange language say. So, Quine argues, when it comes to meaning and interpretation, there is no fact of the matter about who is correct. What is more, the problem of indeterminacy of translation obtains "at home" in English or any language we speak with one another. There are indefinitely many equally good translation manuals for getting from the meanings in my head to those in yours! Accordingly, Quine concluded, an empirical science of meanings or interpretation is impossible.

Moreover, when it comes to the study of other cultures, it will turn out that their ways of thinking, or "conceptual schemes," cannot be radically different from our own, or if they are, we will be unable to detect the difference. For the principle of charity that we require, in order to even try to find the right translation of what they say, requires that we attribute roughly the same beliefs, the same logical principles and perceptual equipment to them that we have!

This is a conclusion disturbing to some cultural anthropologists and others eager to deny privileged status to Western beliefs, especially those that identify the methods of natural science as the sole route to knowledge of the world. These multicultural students of society also want to deny that non-Western values, and especially the religious beliefs of non-Western peoples, are primitive or ignorant. If every normal Homo sapiens has the same set of basic beliefs, the same logic, and the same commonsense epistemology (theory of knowledge), then ultimately all divergent beliefs of people of different cultures will have to be mutually translatable and open to assessment for truth on a single shared standard.

Cultural anthropologists and humanists of various sorts—especially literary scholars—have been particularly unhappy about these sorts of arguments, for two different reasons. The principle of charity homogenizes cultural differences in a way that encourages Western people to conclude that their basic conceptual scheme—the one that gives rise to natural science—is the only correct one. For it is the only one consistent with treating other cultures as coherent. By imposing our conceptual scheme on everyone, no matter what their culture, class, race, or gender, the principle of charity rationalizes a sort of cultural hegemony or imperialism that these scholars have strong political and ideological motives to reject. What is worse, Quine's arguments about meaning also make interpretative social science and literary studies a pointless exercise in which there is no right or wrong. For they show that even after we conclude that we share the same conceptual scheme, the same logic and perceptual beliefs with everyone else, no matter how different their culture, we still cannot select one or even a small number of interpretations of their behavior as better than an indefinite number of others. This makes it impossible to settle debates about the correct interpretation of any cultural phenomena. Without such a possibility, the debates are therefore pointless. Accordingly, much of the debate about the foundations of interpretative social and human science has turned on disputes in the philosophy of language about the nature of meaning.

Against Quine's skepticism about uniquely correct meanings and deconstruction's multiplications of them, there is a powerful argument with an equally French starting point. The argument begins with a claim of René Descartes, the founder of modern epistemology: some things I know for certain, directly, and immediately. I know them without evidence, or they are self-certifying, or the evidence for them must be more powerful than anything that could justify my other doubtable beliefs. Among the things that I know with certainty is my own existence (*Cogito, ergo sum*, "I think, therefore I am"). I cannot doubt my own existence. For the act of doubting requires a doubter—and that would be me. Similarly, at least sometimes when I act, I know why I act. I cannot doubt that my basic actions—raising my arm, for example—stem from desires and beliefs, which give them meaning. My direct awareness of the *phenomenology*, the sensory awareness of what is going on inside of me, guarantees that. The feeling of intelligibility is based on the immediate certainty about why I act, in my own case. And this conviction requires no further evidence, for it cannot be doubted, any more than I can doubt my own existence.

Thus, the argument continues, the reason a proposition like [L] seems like an irrefutable definition is that it is an a priori truth or at least a proposition I know to be true from my own case. I know it by introspection and with far greater assurance than I know any other fact about the world beyond my immediate experience. Of course, whether the other human bodies that make up the social world are just like me "on the inside" is another matter. But if they are, then the intelligibility of their behavior is certified as knowledge by my direct awareness of the truth of [L].

Here we are faced with epistemological rationalism as the final arbiter of method in the social sciences. Meanings provide knowledge, regardless of their predictive limits, because they produce intelligibility, and intelligibility, not prediction, is the mark of knowledge.

Introduction to the Literature

Hegel's works are difficult, whether read in German or English. Accessible treatments of them include two books by Terry Pinkard, *Hegel's Phenomenology: The Sociality of Reason,* and *Hegel: A Biography*. Among other writers on Hegel in English are Robert Pippin, *Hegel's Idealism: The Satisfactions of Self-Consciousness*, and *Hegel's Practical Philosophy: Rational Agency as Ethical Life*, as well as Allen Wood, *Hegel's Ethical Thought.* A more recent work is Stephen Houlgate, *An Introduction to Hegel: Freedom, Truth and History.*

Two excellent works on Marx for philosophy of social science are Jon Elster, *Making Sense of Marx*, and Gerry Cohen, *Karl Marx's Theory of History: A Defence*. Older work but still worth reading are Stephen Lukes, *Marxism and Morality*, and Robert Paul Wolff, *Understanding Marx.*

Accounts of Freud's theory are widely available, but it is best to read Freud himself, for instance, *New Introductory Lectures on Psychoanalysis*. A. Grunbaum, *The Foundations of Psychoanalysis*, is a philosophical critique of Freud's theory, with special reference to the critical theorists' reinterpretation of it. A more recent interpretation of Freud's research program is Patricia Kitcher, *Freud's Dream: A Complete Interdisciplinary Science of Mind.*

Much of Foucault has been translated into English. Although less difficult than Hegel, Foucault's prose is not easy. Gary Gutting's *Michel Foucault's Archaeology of Scientific Reason* is a good place to begin. Gutting's anthology, *The Cambridge Companion to Foucault,* is testimony to Foucault's continuing influence. Bourdieu's work has now been translated into English as well, and has an increasing influence among sociologists of culture. M. Grenfell's anthology of papers, *Pierre Bourdieu: Key Concepts,* is a good starting point, along with J. F. Lane, *Pierre Bourdieu: A Critical Introduction*.

Among critical theorists, the most influential figure for social science is J. Habermas. One of his important works is *Knowledge and Human Interests*. T. McCarthy, *The Critical Theory of Jurgen Habermas*, is recommended as an excellent introduction to his thought. See also R. Geuss, *The Idea of Critical Theory*, a particularly lucid exposition of the theory. McCarthy's *Ideals and Illusions* is a more recent text on the subject.

The general problem of reflexive predictions in social science is examined in papers by Buck and Grunbaum reprinted in L. I. Krimerman's anthology, *The Nature and Scope of Social Science*, and G. D. Romanos's "Reflexive Predictions" in Martin and McIntyre.

Quine's attacks on meaning and his argument for the indeterminacy of translation can be found in *Word and Object*, arguably the most important work of philosophy in the United States during the twentieth century.

Martin Hollis, in *The Philosophy of Social Science*, explores the prospects for reconciling interpretation and behaviorism under the distinction between explanation and understanding. Hollis also treats the once-lively debate about whether there are alternative incompatible conceptual schemes, which cultural relativism requires.

Students seeking further discussion of "deconstruction" do well to begin and end their inquiries with "The Word Turned Upside Down" by John Searle, a review of *On Deconstruction: Theory and Criticism after Structuralism* by Jonathan Culler in the *New York Review of Books.*

CHAPTER NINE

Holism and Antireductionism in Sociology and Psychology

Among social scientists there is a long tradition of insisting that the distinctive subject matter of their sciences are facts that are not psychology, whether matters of prediction or interpretation. Instead their sciences explain autonomous social facts by other social facts, and neither is reducible to more basic facts about individual people or anything else. The same type of argument is advanced by psychologists to deny the relevance of more basic sciences, such as neuroscience, to its research agenda. In this chapter we examine the structurally similar arguments of antireductionist social scientists and psychologists.

HOLISM AND ANTIREDUCTIONISM IN SOCIOLOGY AND PSYCHOLOGY

Not all social science is devoted to the psychological explanation of individual human action, nor to explaining human institutions as dependent on individual psychological processes. Indeed, the tradition of European philosophical anthropology examined in the last chapter holds that it is broad social forces that operate on individual psychologies to constrain, structure, or even produce individual action. Even among those who embrace an empirical, naturalistic approach to human affairs quite different from the European tradition, there are sociologists, anthropologists, economists, and students of politics who hold that what is characteristic about social science is that it deals with a range of facts about people, facts that have little to do with the psychological factors explaining individual human action. The facts are about human social institutions, like families or businesses, and about large aggregations of people, like social classes or religious groups, or even about whole societies, economies, or cultures. Following a tradition dating back to Emile Durkheim, one of the founders of sociology, let us call these facts "social facts." That social facts are not the same as the sum of the psychological factors explaining individual agents, their behavior, and its causes is crucial to the claim that special social facts exist.

SOCIAL FACTS

If there are such social facts, then there might be a way around all the problems of intentionality. A theory that explains and predicts the relations among social facts, without reference to psychological facts, need not concern itself with problems of intentionality that psychological factors raise. A theory of social facts may make no use of an explanatory principle like [L] or any other psychological laws. Moreover, the existence of a range of such facts may provide a better explanation for the predictive weakness of social sciences: perhaps the problem is that we have been attempting to construct a theory about the wrong sort of subjects—individual people, instead of social groups. If we change our perspective, we will discover an improving science staring us in the face. Consider a parallel from physics: imagine trying to frame a theory about the thermodynamic properties of a gas—including temperature, pressure, volume—by focusing on the behavior of the molecules that make it up. Even if we could observe the molecules, we would see nothing but a buzz of random motions, nothing that could help us understand the gas composed of them. The proper way to proceed in thermodynamics is to search for macroscopic regularities among observable features of the gas. Only later may we frame hypotheses about individual molecules that might explain the laws of thermodynamics.

The same may be true for the social sciences. The behavior of the individual looks random. We can't seem to frame any laws about it. But what if we examine aggregations of individuals? Then we might—indeed, it has long been argued, we will—find laws about these aggregations of individuals. Having found them, moreover, we may be able to turn back to discover hypotheses about individuals that explain these macrosocial laws. And even if, as some hold, there are no such explanatory hypotheses, at least we will have discovered important laws of macrosocial phenomena.

Advocates of the possibility or the actuality of social facts are called *holists*. Holism is often associated with the doctrine that the whole is more than the sum of its parts: society is more than the individuals who make it up—hence the existence of independent social facts.

There are, however, philosophers and social scientists ready to produce arguments that there is no such range of irreducible social facts to be found or explained. These opponents of social facts are known as *methodological individualists*, or *individualists* for short. They stand ready to demonstrate that any example given of a social fact is best described or fully explained in terms of the behavior of individuals, or else it is not a fact of any kind, but a figment of the holist's imagination or research program. The modifier *methodological* is attached to the name of proponents of this view in order to reflect the fact that they adopt a methodological dictum always to search for individual descriptions and explanations to substitute for or explain holistic ones.

It will be surprising to some social scientists that there could be a philosophical question about whether there are social facts independent of individual ones. For such independent facts seem the sum and substance of many of the social disciplines. In the view of many social scientists, that such facts obtain is surely a substantial matter of fact, not a question to be decided by considerations from philosophy. However, there is a long tradition (especially but not exclusively) among naturalists in social science and philosophers of science of doubting whether there can be any such facts not reducible to facts about individuals.

In part, this skepticism stems from a suspicion of theoretical entities—things, forces, or features that are not directly observable by the scientist. Some physicists and philosophers of physics have raised doubts about the existence of microphysical entities, such as electrons, quarks, and photons. Because we cannot observe them, we cannot justify our theories about them as any more than useful fictions. We can only observe medium-sized objects and their properties. Therefore, these are the only sorts of objects and properties we have reason to believe really exist. Similarly, we can observe only individual agents, so that claims about things very much "larger" than people are just as suspect as claims about things very much smaller than people. There is another reason naturalists are dubious about social facts. The very idea of a range of facts distinct from ones about individuals seems intrinsically mysterious and unempirical to them. For such facts smack of "organicism"—the doctrine that, like an organism, the social whole is something above and beyond its parts and their relations to one another. Similarly, they reject the idea that stems from Hegel that above and beyond the individual consciousness that governs behavior there is a "collective consciousness" of the society as a whole. Even holists who reject both organicism and the doctrine of a collective consciousness are sometimes accused of commitment to them willy-nilly.

Holism is also controversial because it is connected to another feature of social science that raises serious methodological questions—"functionalism," the strategy of identifying and explaining features of society in terms of the functions, the purposes they serve, not for individuals, but for the society as a whole. This strategy is widespread in social science. And as we shall see, holism provides a natural motivation for it. However, functionalism has itself been a problematical strategy, in part for reasons we have already examined in the discussion of behaviorism and teleology (Chapter 5, "Causation and Purpose").

Part of the argument for holism derives from the motive of showing sociology to be an "autonomous science"—one distinct from and independent of psychology. Arguments for the autonomy of a discipline have usually been made at the time of its first separation from another established discipline. Thus, Durkheim argued strongly for the separate existence of sociology and its distinction from social philosophy and individual psychology. Arguments about autonomy also arise when an established discipline, theory, or method appears to be threatened with elimination or preemption by a more powerful or more basic one. Thus, some psychologists hope to show that psychology is independent of apparently more fundamental theories and methods in the life sciences, like those of neuroscience. And even more social scientists reject the dependence of the human sciences on any part of biology, especially Darwinian theories of evolution.

Opponents of the autonomy of a discipline or theory often hold that it is "reducible" to some more fundamental theory or theories. Recall the brief sketch of logical positivism in Chapter 2. According to positivists, the history of science is the history of the subsumption of older and narrower theories by newer and broader ones as science advances. More specifically, in theory reduction the terminology of the narrower theory is defined in the terms of the broader theory, and its laws are derived as special cases from the more general laws of the broader theory. Thus, for example, important parts of chemistry lost their autonomy from physics when the balanced equations of chemical reactions came to be derived from the theory of atomic structure in physics.

Methodological individualists argue that if there are social facts, and generalizations about them, then the theories about these facts can be reduced to one about individuals. That is, the terms naming social facts can be cashed in for terms that describe individuals and their properties, and the laws about social facts can be derived from psychological laws like [L]. Similarly, proponents of the reduction of psychology to neuroscience argue that its intentional vocabulary and its generalizations can be reduced to descriptions of brain states and laws about their relations.

The arguments for the autonomy of sociology and psychology are so similar that we will examine them together and draw conclusions about both subjects. In Chapter 10 we turn to the connection between holism and functionalism and the methodological individualist's critique of both. And in Chapters 10 and 11 the strategy of methodological individualism, particularly in economics and in sociobiology, is examined further.

HOLISM AND HUMAN ACTION

How is the existence of social facts to be established? One argument for holism is narrowly philosophical and turns on the analysis of the terms we use to describe individual human actions. A second argument is more clearly factual and, therefore, more convincing to social scientists than the philosophical one. Before turning to it, we examine the philosopher's argument, with the warning that it, like most philosophical arguments, is pursued at a very high level of abstraction.

Consider the terminology employed in intentional explanations. Merely by reflecting on notions like language, meaning, rules, roles, and institutions, we can deduce the existence of such social facts. Describing something as an action and explaining it in terms of beliefs and desires seem already to commit us to social facts. Consider the example of my cashing a check. How is such behavior explained? In order not to miss aspects of this explanation suppressed because of their banality and obviousness, imagine explaining this behavior to someone who does not understand the concept of money. To do that, I need to explain the significance of my action to me and the teller, and that brings in the entire monetary and banking system. The teller and I are operating under rules—enforced by society—that give the exchange its meaning. But the rules are unintelligible except against the background of rules with compliance conditions, that is, institutions of persuasion and enforcement. My actions are explained in terms of the status of the teller as a teller, independent of any other facts about him. His behavior as a teller can be described only by reference to other roles—the cashier, the bank manager, and so on. We cannot break out of this circle of institutions into descriptions of mere behavior. An account of the beliefs and desires that give the meaning of each of our actions must make reference to our roles and the rules governing them. Since we cannot characterize these beliefs and desires without reference to concepts like *teller* and *customer* that give their content, reference to social facts is unavoidable in individual explanations. If reference to social facts is unavoidable in describing individual facts, they can hardly be composed of or dependent on them. Social facts must have a separate and distinct existence. So the argument goes.

The methodological individualist's traditional counter to this argument was an objection from considerations of testability. Individualists held, quite rightly, that the only test of a statement that purports to refer to social facts is to be found in observations about individuals. Therefore, such statements had ultimately to be translatable, at least in principle, to claims about individuals. Any residue in such translation was a meaningless leftover. The trouble with this argument was that it proves too much. Suppose we employ the same standard in, say, physics, and demand that all of its theoretical claims must also be translatable without residue into statements about observations. That cannot be done short of depriving physical theory of its explanatory power. If "electron" must be translated into observational terms, we can no longer explain its observational effects such as lightning by appeal to the behavior of electrons. For now what explains the lightning turns out to be a disguised statement about lightning itself.

What this point shows, however, is that statements that transcend observation are to be judged on their explanatory power, not their testability or meaning. Therefore, the argument that the meaning of our very descriptions of human actions presuppose social facts does not carry much weight. For such statements merely describe phenomena that need to be explained. Perhaps the best explanations will involve redescription of the events in terms that do not presuppose social facts. A description of the sun as "setting" seems to commit us to the geocentric hypothesis: the sun goes around the earth. An explanation of the sunset, however, makes no such supposition. It begins by redescribing the sunset as an earth-turn. Similarly, whether there are social facts or not cannot rest on descriptions we employ to identify individual actions. It must rest on the argument that the best explanation of individual actions presupposes social facts.

Therefore, holism requires more than simply showing that our descriptions of individual actions presuppose social facts. Any argument for the existence of social facts must rest entirely on the adequacy of an explanatory theory of human actions that requires such facts. The explanatory theory that adverts to social facts must provide the best explanation for the character and content of the beliefs and desires that result in actions.

But the methodological individualist will not accept as adequate any such explanatory theory appealing to social facts. The individualist holds there is a crucial disanalogy between the unobservable, theoretical facts of physics and the unobservable, social facts of the holist. For the existence and interaction of individuals are not only necessary for the existence of society and of social facts but are also sufficient for their existence. Both holists and individuals agree that people are necessary for the existence of society. Individualists argue that, in principle, claims about social facts should be entirely explained by truths about these individuals and their relations to other individual people. This is because the existence of multiple individuals is sufficient for the existence of society and social facts. If all we need to produce social facts is for people to exist and behave in certain ways, then how can social facts consist of more than people's behavior?

If indeed people's existence and behavior in certain ways are all we need to produce social facts, then the comparison between theoretical facts of physics and social facts breaks down. The truth of claims about the observable phenomena of physics is necessary for the truth of statements about the theoretical facts that physics postulates. That's how we test the theoretical ones. But such confirming observational predictions are not sufficient for the truth of theoretical claims. Otherwise, we could translate statements of physical theory into sets of statements about observations, because translation requires just the knowledge of necessary and sufficient conditions for the term translated. Then, with theoretical claims equated to descriptions of observations, the power of statements, say, about electrons, to explain observations of, say, lightning, would be a real mystery. By parity of reasoning, if aggregating individual behavior is sufficient for bringing about social facts, then the social facts cannot explain the individual ones because individual behavior is already necessary for it.

Now, one way holists can restore the parallel is to claim that the whole is somehow different from the sum of its parts and their relations to one another: somehow, putting people together creates new things that, together with people's behavior, constitute society and make for social facts. These new things are the theoretical entities of sociology. Just aggregating or piling up individual actions isn't after all sufficient for the existence of social facts. Moreover, we must appeal to sociology's theoretical entities to explain the behavior of individuals in society. Few philosophers or social scientists will want to take this idea seriously, except as a last resort. For in the end, this hypothesis is not so much an explanation of anything as an admission that something can't be explained. How the whole could be different in kind from the sum of its parts and their relations is just a mystery.

But consider the alternative that holism faces. It is the view that there are two sets of facts, one about society, and the other about individuals. The first set is wholly dependent on the second and yet not identical with it. The dependence is not accidental; it's at least causal and perhaps logical. Without people, no social facts. With people, social facts. Moreover, these social facts, whose existence depends only on the existence and interaction of people and on nothing else, are an essential part of the explanation of people's behavior. It's not just that individuals' subjective and perhaps false beliefs about social facts explain their individual behavior. On the holist's view, the social facts are needed to explain their beliefs, true or false. Perhaps the hypothesis about wholes being greater than the sum of their parts is more attractive than first supposed. For it seems no more mysterious than the complete dependence of social facts on psychological ones, even though the dependent social ones explain the independent psychological ones.

In the end the argument for holism from the terms in which we describe action is not very appealing. And if the intentional explanation of action really requires social facts, as this philosophical argument holds, then holism—with its mysteries about wholes being different from the sum of their parts—will also have to bear the methodological problems of [L] and intentional explanations generally. As such, it would fail to provide a potential way out of the dilemmas of naturalism and interpretationalism. So, it's pretty clear that holism needs an argument that is both more convincing than this philosophical one and independent of the problems of intentional social science. As we shall now see, there are more powerful arguments for holism, arguments that, in fact, undermine intentional theories of human action.

THE AUTONOMY OF SOCIOLOGY

Among speculative philosophers and in various religious traditions, holism is an ancient doctrine. Hegel's thesis that the history of the world is the story of the self-development of the world mind or spirit is the best example of this tradition. But holism became an issue of importance to social science with the work of Durkheim in the late nineteenth century. Durkheim did more to establish sociology as an independent discipline than anyone else. However, Durkheim was as hardheaded an empiricist about methodology in social science as one is likely to find. He was quite explicit in claiming that the methods of sociology must be the same as those of the natural sciences. As to the subject matter of sociology, he held that it is quite distinct and different from the subject matter of any other discipline.

In fact, the existence of a range of social facts was Durkheim's most powerful argument for a separate science of sociology, at a time when it was struggling for autonomy from psychology (and social philosophy, for that matter). In the 1890s, even psychology was not yet viewed as a distinct discipline, independent from philosophy. And what better argument for establishing a discipline's autonomy can there be than showing there are facts that no other discipline even takes note of, let alone can explain?

The term *social facts* was coined in Durkheim's methodological treatise *The Rules of the Sociological Method*, but the most powerful argument for them was not methodological. It was factual. In *Suicide*, Durkheim uncovered some startling statistics. Among the best known nowadays is that the number of suicides per 100,000 differs radically for Catholics and Protestants, even when we control for all other reasonable factors. Catholics have a much lower suicide rate. The numbers also differ spectacularly and significantly between men and women, married and single, army officers and conscripts, the newly wealthy and longtime impoverished, summer and winter, during meals and between meals. These statistical differences are large enough and persistent enough to ground regularities that require explanation.

Durkheim noted that the reasonable thing to do is to examine the reasons for suicide and see how they vary between the differing classes, Protestant and Catholic, officer and conscript, and so on. For suicide is an action, and accordingly is the result of desires and beliefs. In fact, coroners had already been recording the reasons for suicide throughout the nineteenth century, citing the presumptive causes of suicide under headings like poverty, family troubles, debauchery, physical pain, love, and jealousy. Of course, there were errors in coroners' reports owing to the difficulty of correctly identifying the reasons from case to case. But under the reasonable assumption that errors will cancel each other out, what did Durkheim find? Well, in certain parts of Europe, the suicide rate rose 100 percent between 1856 and 1878—and not gradually either. Rather, it remained at the lower level for some years, then suddenly jumped to the higher level and remained there. But now Durkheim examined the coroners' reports: he found that the proportion of each of the "presumptive causes" of suicide—the reasons—illness, poverty, jealousy, and so on, remained almost exactly the same between 1856 and 1878, while the rate of suicides per 100,000 population doubled everywhere. That meant that either the incidence of each of these reasons for suicide increased in exactly the same proportion or they were not the causes at all.

Now, recall the methodological maxim: same cause, same effect. This principle tells us that like effects must have like causes and unlike effects must have unlike causes. In 1856 and in 1878 the effects—the number of suicides per 100,000—were unlike. There was a 100 percent rise. So, whatever caused the rate in 1856 cannot also be identified as the cause in 1878. Since the reasons cited in the coroners' reports remained the same, in proportion, they cannot have been the causes. Therefore, Durkheim concluded that even if each individual suicide is caused by a psychological factor—the suicide victim's reasons for committing suicide—the change in the rate per 100,000 cannot be caused by such facts. The statistical fact is a social fact, Durkheim held, and cannot be explained by psychological facts. To explain this social fact, we must seek other social facts that cause it and that also cause the psychological facts. For the stability of reasons for suicide over time must itself be explained by forces external to and independent of the individual.

Here is an argument for the existence of social facts independent of a theory that hypothesizes them. The argument relies only on empirical facts and a methodological principle everyone shares: same cause, same effect. And once we grant, on the basis of an empirical argument, the existence of such facts, there is nothing methodologically suspicious about hypothesizing further social facts to explain the ones whose existence we have already proved without a question-begging appeal to a theory. But that is exactly what Durkheim did.

Durkheim's theory is a paradigm of twentieth-century sociological theorizing. The theory treats the suicide rate as a social indicator—a measure of the health of the society as a whole. Beyond some baseline level of suicides per 100,000, which we can determine by comparing time-series data, a rise in the suicide rate represents something gone wrong in the society. The members of a society are governed by social forces that exist independent of them. The behavior of individuals is determined by norms of conduct of which we are not aware. These norms are imposed on us by social institutions, which determine the degree of social integration of a society's members. Thus, suicide is lower among Catholics because they are bound more closely to one another through the guidance of the Church than are Protestants. And suicide is higher among army officers than conscripts because officers are bound more tightly to their units than conscripts are and officers sacrifice themselves to "the good of the service" if required. The newly rich person, who is unused to luxury and whose means now exceed his needs, is made normless and more inclined to suicide than perpetually poor folk, whose social norms remain undisturbed. In fact, Durkheim identified three different sorts of suicides in terms of three distinct social causes: egoistic suicide, resulting from too little social integration; altruistic suicide, resulting from too much of it; and anomic suicide, resulting from great and rapid changes in the degree of social integration that leave agents normless and disoriented.

We can explain each of the social facts about differences in suicide rates by a social fact about differences in the degree of social integration. There is an optimal level of social integration at which the baseline level of suicide is maintained. But optimal for whom and for what? Durkheim's answer was that a certain level of social integration is optimal for the society as a whole, for its well-being and survival. Each of the institutions of society has a function to fulfill, and its behavior is explained by this function. Thus, marriage, religious organizations, family structure, legal and business institutions, political organizations, in fact, most things of interest to the sociologist, have a function in the operation of society. When these institutions are oppressively overbearing, the society functions poorly, as manifested by a high suicide rate and by other social ills as well. When institutions are not sufficiently powerful in their effect on individual behavior, the same symptoms of social disorder are manifest. Durkheim viewed society as a vast organism, with a complex organization of interactive components. Therefore, the function of a social institution is not to do something for individuals directly but to do something for the society composed out of them. This is holism with a vengeance—and with no doubt about the existence of entities above and beyond individuals.

Durkheim was not embarrassed about the apparent commitment of his theory to a doctrine of "organic wholes" and its concomitant thesis that the whole is different from its parts and their relations to one another. There were two reasons for that: First, along with other late nineteenth-century writers about method, he supposed those principles to be substantiated in the relation of physics to chemistry and in the relation of both disciplines to physiology. That is, chemical phenomena could not then be explained wholly in terms of physical theory, nor could physiology be explained fully in terms of chemistry. Therefore, chemical properties of things were treated as distinct from the physical properties of their constituents. Similarly, metabolic and other physiological processes could not be chemical or physical explanations. So for Durkheim, holism was a widely accepted scientific stance.

Durkheim seems to have held that society, along with individuals, is made up of an *âme collective*—which is usually translated as "group mind," though we should beware of literal translations. It is unclear how much Durkheim really required this notion in order to expound his theory. What is clear is that neither it nor the social regularities it was called upon to explain was, in Durkheim's view, reducible to psychological processes. Therefore, sociology had to be an independent and autonomous discipline. It had its own facts and its own laws. What more could a science want? Its own distinct method? No. Durkheim held to the thesis that logical positivists later called the "unity of science"—that disciplines might be distinguished by subject matter, but not by method. For scientific method is determined by the requirements of knowledge, which are the same everywhere.

In fact, according to Durkheim, psychology must surrender some of its domain to sociology. That is because his proof that social facts exist is also an argument that at least some individual behavior that we suppose is determined by psychological factors is not, in fact, really caused by them. Durkheim recognized that social forces must work through individuals. So he held that such facts constrain and direct individual behavior, even when the behavior seems entirely "voluntary," in his words. And Durkheim had an argument for this conclusion: if there are social forces that determine the suicide rate, they can do so only by determining individual suicides. What the psychologist identifies in the mental states of individual suicide victims were accepted by Durkheim as sometimes being the real intermediate links in a causal chain from the social forces to the act of suicide. At other times Durkheim seemed to say that these mental states are mere by-products, produced along with suicide by social forces.

Here, then, is a more powerful argument for holism, and not just about the social suicide rate. For Durkheim and his sociological successors have employed similar arguments to assert the existence of a wide range of such facts. Many of these successors, especially in the European tradition of philosophical anthropology, took great inspiration from Durkheim's quantitative and data-driven argument to hold that there had to be social facts distinct from psychological ones; that the relationships among these facts constituted an autonomous set of social structures; and that, as Durkheim discovered, these facts and the structures in which they are embedded determine individual conduct unconsciously and/or consciously. Figures such as Foucault, Bourdieu, and Habermas all took inspiration from Durkheim, though none explicitly identified themselves as a disciple or protégé.

But Durkheim's influence in empirical sociology is at least as great as it was on the more speculative European tradition. If Durkheim's argument is sound, there are more than enough social facts to keep an entire scientific discipline busy. This type of argument is one that claims to be well within the naturalistic camp, at least as regards scientific aims and methods. It claims to be an argument no different from those that stand behind the autonomy of biology from chemistry or physics. No one can doubt that there are biological facts, whose existence we recognize without having already bought into biological theory. So, too, are there sociological ones, whose existence can be established without a question-begging appeal to a theory that already assumes them. But once we recognize the existence of some social facts that cannot be explained by psychology, there is another argument for such facts. For the explanatory power of the theory of autonomous social factors that best explains them provides further evidence for the truth of holism.

HOLISM AND REDUCTIONISM IN PSYCHOLOGY AND SOCIOLOGY

There is, however, at least one loose thread in this argument. It's the relation between autonomous sociological facts and the psychological ones from which they are supposed to be autonomous. Consider the psychological facts about individual suicides, for example. How can we decide whether psychological factors are causal links to suicide or just by-products of some other suicide cause?

The question is important for Durkheim's claims about the autonomy of sociology and the existence of social facts. In fact, it is a crucial issue for any argument that justifies the autonomy of a discipline on the existence of distinct entities and distinct laws about them. If it is shown that the occurrence of social facts can be explained by laws of another, more fundamental discipline, then the argument for the autonomy of the discipline seems seriously weakened. For then it may be claimed that the facts the discipline deals with are not autonomous but can be described and/or explained by the more fundamental discipline. This argument has always seemed especially threatening to sociologists and anthropologists. For without an argument for holism, their disciplines seem in danger of being swallowed up by psychology. Without holism, what else is there to society but people, whose behavior is the business of psychology to explain? Sociologists and anthropologists therefore frequently quote Durkheim's exhortation, "Whenever a social phenomenon is directly explained by a psychological phenomenon, we may be sure the explanation is false." We need to consider whether this thesis is defensible and what sort of autonomy from psychology it is that sociology needs or can secure.

Suppose that psychological facts are causes of suicide in individual cases. Then presumably there will be generalizations linking these factors to suicide, perhaps even intentional generalizations of [L]'s form. If there are such generalizations, then by simple arithmetic aggregation of the explanations of individual suicides, the generalizations should also explain the aggregate fact of the number of suicides per 100,000. If Durkheim is right, working upward from the psychological causes of suicide should bring us to the social facts—too much or too little social integration—that determine these psychological causes. In effect, psychological factors would then be part of the explanation of why social facts obtain. Psychological laws would underlie and help explain sociological laws linking the degree of social integration to the suicide rate. But this result threatens the autonomy of sociology. For it makes it look as if sociology is "reducible" to psychology. All we would need is to show that social integration is itself the result of psychological factors or the behavior of people toward one another. Then it will turn out that psychological facts about people cause them to treat others in a way that leads the others to commit suicide.

The alternative to this reduction of the sociological is to treat psychological factors as by-products. We must claim that they are "epiphenomena," which have no causal role in suicide, that psychological factors are side effects, along with suicide, of purely social forces. But this claim flies in the face of a causal principle that no empiricist like Durkheim could ignore: the principle that there is no "action [that is, causation] at a distance." One of the legacies of the success of mechanical explanations in science is the doctrine that one change cannot cause another unless they are in spatial and temporal contact or unless there is a chain of such contacts between them. Now, for a change in the degree of social integration to cause a change in the suicide rate, there must be such a causal chain, and it must pass through people. Unless it can do so without passing through their thoughts, psychological processes cannot be mere by-products of suicide's social causes.

The same argument seems available for any generalization connecting one social fact with another. The causal chain must pass through individuals, and this threatens the autonomy of disciplines that deal with social facts. Furthermore, showing the dependence of social laws on psychological processes may lead to the conclusion that the social facts are ultimately psychological too.

This conclusion, in fact, mirrors the logical positivist image of the nature of scientific progress and the structure of scientific knowledge, already described. The image seriously threatens arguments for the autonomy of distinct social sciences, especially ones without well-established records of success in the discovery of laws. The image, as we saw, is unabashedly reductionistic. It claims, first of all, that the history of scientific progress is a history of reductions of narrower theories to broader ones. Once a science discovers its first improvable generalizations, progress comes in the formulation of deeper laws and theories that not only explain the initial generalizations but also improve on their accuracy. Thus, Kepler's laws of planetary motion and Galileo's laws of terrestrial motion were the break with Aristotle that produced modern physics. It took Newton to show that both were derivable from a single set of laws, one of profound economy and simplicity, to which many other regularities were "reduced" over the next three centuries. That is, these regularities were shown to be special cases of Newton's laws or deducible from them when we added certain assumptions about the mathematical values of parameters and constants and so on. Newton's laws enabled us not only to systematize disparate regularities but also to improve them, explain their exceptions, and show us what further forces need to be taken into account in order to improve our regularities. Scientific progress eventually led to the reduction of Newtonian mechanics, to still more fundamental principles that explain it and its exceptions: the theories of relativity and quantum mechanics. We can deduce Newton's laws from these theories by adding the false assumptions (embedded in Newton's theory) that the speed of light is infinite and that energy comes in a continuum of values, instead of discrete quantities—quanta.

In addition to reducing Newtonian mechanics to more fundamental theory, modern science can reduce thermodynamics, electromagnetism, and large parts of chemistry to fundamental physical laws. In fact, it is claimed that important parts of biology—such as genetics, enzymology, and parts of physiology—can be reduced to chemistry. That suggests not just that science progresses by reduction but that the edifice of scientific theories is reductive as well. Thus, chemistry is reducible to physics, and increasingly, biology has been shown to be reducible to chemistry. But where do the social and behavioral sciences fit into this picture?

The empiricist and postpositivist picture is a reductionist and antiholist one. It tells us that biological systems are nothing but chemical systems and chemical systems nothing but physical systems. So, psychological systems—organisms with minds—must themselves be biological and, ultimately, just chemical or physical systems. Social systems—groups of individuals—must ultimately be psychological ones. If there are psychological laws, they should be derivable from biological ones; if there are sociological laws, they should be derivable from psychological ones. This reductionistic view of the history and structure of science has had more proponents in the sciences than in philosophy.

The methodological moral reductionist draw seems twofold. First, propositions that are not explainable by reduction to laws of a more fundamental discipline are not laws but are either falsehoods or local regularities that describe the results of the operation of more basic laws on initial conditions. Second, any discipline that has not yet secured laws is unlikely to do so, unless it follows the guidance of methods that have secured laws in other disciplines and employs descriptive language common to these successful disciplines.

This view has profound ramifications for psychology that we must address before returning to its bearing on holistic arguments for the autonomy of sociology. As we have seen in Chapter 4 ("Intentionality"), intentional psychology seems irreducible to neurophysiology. Neuroscientific theory does accord brain states "content" in the way that the intensional descriptions of folk or much scientific psychology do. Substitution of equivalent descriptions in any sentence of neuroscience will not change a truth to a falsity, or vice versa. The same is true of the rest of science. That means that the intentional descriptions psychology provides of our mental states can't be equated with neurological descriptions of brain states. But just such equations are necessary conditions for reduction. To see the problem, compare a successful reduction in physics. We can deduce the ideal gas law, PV = nRT, from theory about molecules in part because we can identify the temperature of a gas with properties of molecules that make it up. Reducing thermodynamics to molecular mechanics hinges on the fact that the temperature of a gas is equal to the mean kinetic energy of the molecules that compose it. A parallel kind of equivalence between belief, for example, and any description of brain states is just not in the cards—it's logically precluded by the intensionality of mental states that neural states (along with all other physical states) lack.

The conclusion a few philosophers, psychologists, and social scientists draw is that intentional psychology is a dead end, that there are no laws of intentional psychology. For any such laws would not be reducible to the rest of science, and that's impossible. Why? Because intensionality blocks the reduction and thus the unification of psychology with the rest of science.

These psychologists and philosophers insist that all our scientific theories should eventually be interconnected and arranged in a hierarchy from the most fundamental to the most derivative, and the derivations must be deductive. This requirement will explain why no one has found any laws in psychology. It will explain why [L] is so close to vacuous and has so little predictive content and, most of all, why it hasn't undergone any improvement in all of recorded history. [L] can't be improved because improvement requires being linked to a broader reducing theory, just what is impossible for intensional statements. And the problem is the descriptive terminology we have always employed in attempting hypotheses in psychology. Intentional concepts don't link up neatly to the rest of science because they don't "carve nature at the joints."

Recall the point about the concept of fish in Chapter 2 ("The Empiricist's Diagnosis"). Any attempt to frame generalizations about how fish breathe, and to improve these generalizations or explain them, will be frustrated. There is no one way fish breathe; *fish* is not a "natural-kind" term. Though its origins in ordinary thought are clear, the category "fish" has no place in biological science, because it cannot be linked up in laws with other general categories. So, biologists give up the ordinary term *fish*, meaning "aquatic animal," and either redefine it or break up the category of fish into several homogeneous categories to which they assign Latin names.

Psychologists who embrace the view that intentional terms don't name natural kinds have acted similarly. They have turned their attention to neuroscience or artificial intelligence or have tried to make a go of one or another version of behaviorism.

This argument that intentional theories are dead ends has several implications for the rest of social science. First, other social disciplines, to the degree that they are intentional, have no more prospects of reduction than psychology has, and for the same reasons. They will not be reducible to extensional science. Moreover, on this view, the failure to have identified laws and theories reflecting the continuing derivation and improvement characteristic of science is explained by the intentional character of psychological theory.

The second implication of surrendering intentionality is more favorable to sociology. If intentional psychology is a will-o'-the-wisp, then of course there is no reason to demand that sociological generalizations be explained by the intentional psychology of individual behavior. A macrosociology irreducible to intentional psychology is untouched by its problems. However, macrosociology must, in the reductionistic approach, link up with the rest of science somewhere, either through a nonintensional psychology or perhaps through some direct connection to biology. Either way, in this argument, sociology will turn out to be reducible to, not autonomous from, the rest of science. This is of course a heads-I-win-tails-you-lose argument against holism and autonomy. If sociology is not reducible, then it's a scientific dead end, like intentional psychology. And if it is reducible, it is not autonomous.

The argument, however, has several weaknesses. To begin with, there is its picture of the history of science as cumulative progress by successive reduction and as a deductive hierarchy of scientific theories. If the picture ever was an uncontroversial account of the history and present status of scientific theorizing, it is no longer. Both historians of science and some philosophers have repudiated it as a simpleminded reconstruction of the much more complex history and structure of science. First, there have been detailed attempts to show that the deductive relation claimed to hold between successive theories in physics does not obtain. Nor do such relations obtain, it is claimed, between theories in different sciences. Biological theory is held to be autonomous from chemistry, even at the level of molecular genetics, but especially between evolutionary theory and the rest of natural science.

Let's consider the least radical of these objections to the reductionist's picture, the notion that the structure of science is not a deductive hierarchy of more derivative and more fundamental theories. Biology is certainly a respectable science, one that adopts to a large extent the causal methods of physical science and that searches for laws to explain its phenomena. Moreover, it has had some important theoretical successes: evolutionary theory, population genetics, molecular biology, to name the most imposing. Yet, biological theory is not (as yet) reducible to chemistry or physics. Even if it eventually is reduced to physical science, no one will dare to deny that biology is a separate science or to deny the reality of biological organisms.

Sociologists and intentional psychologists can take considerable comfort in these facts. For they suggest that irreducibility is not a symptom of pseudoscience or sterility and frustration. Moreover, even if sociology were reducible to psychology, and/or psychology were reducible to neuroscience, it would not follow automatically that there are no sociological facts or no psychological facts after all.

Actually, the question of whether biology is reducible to physical science may be instructive for the holism/autonomy question. For one thing, it suggests that the really interesting issue is not whether there are social facts, but whether there is a distinctive discipline couched in the language of such facts. Almost all biologists are prepared to admit that their research subjects, whether the species, populations, organisms, organs, tissues, cells, or the macromolecule, are "nothing but" physical matter, albeit organized in distinctive ways. No biologist thinks each of these levels of description refers to a distinct and different entity greater than the sum of its parts. Rather, biologists are interested in whether interesting and useful generalizations can be discovered at each of the various levels of organization they observe (organisms and organs) or theorize about (the species or macromolecule). One way of saying that there are such generalizations at a given level—such as the level of species or macromolecule—is to say that there are facts about the species or macromolecule distinct from facts about individual organisms. And such a claim need have no mysterious metaphysical or ontological connotations.

Similarly, the autonomist may argue, the question of whether there are social facts is the question of whether there are interesting generalizations couched in language that purports to refer to such facts. The rest is "mere" philosophy. Let the philosophers fight about whether the terms in a well-confirmed law refer to real objects. Like the claims of biology, sociological laws must ultimately be explained by psychological ones or whatever theory best explains individual human behavior. But that doesn't mean there are no interesting generalizations about social facts. There may still be social generalizations that can be used to explain and predict social phenomena, and some cases of individual behavior, for that matter.

This is not a view Durkheim would have been comfortable with, perhaps, while fighting for sociology's life. But now that its life is no longer threatened, we may relax and adopt it as the cognitive content of the claim that there are autonomous social facts. This is a view we might describe as "methodological" or "instrumental" holism, according to which the autonomy of a discipline hinges, not on whether a special range of facts exists, but on whether the discipline can come up with interesting generalizations.

As an epistemological or methodological thesis, antireductionism and the autonomy of biology from physics, or the macrosocial sciences from psychology, or for that matter psychology from neuroscience, is probably not controversial. But it may also not be strong enough to support the sort of autonomy of the social sciences from psychology and from the natural sciences, especially biology. And that the generalizations of the social sciences are both interesting and apparently irreducible to psychology, and that psychological regularities are not in fact reducible to neurobiology, has emboldened philosophers of social science and philosophers of psychology to try to construct arguments that these regularities could not be reduced, even were we to know everything, even if the social and behavioral sciences were complete. If these writers are correct, the autonomy of higher-level sciences from lower-level ones will not just be epistemological or methodological. It will be a metaphysical difference in kind between the domains of these disciplines. This argument has become important in recent years and needs to be explored.

AUTONOMY AND SUPERVENIENCE

The argument for the stronger metaphysical autonomy of any social science from other more fundamental sciences begins with a distinction that will be important in the next chapter as well. It is the distinction between functional kinds or types of things, states, and processes, and structural types. This is not a hard and fast distinction and it is easy to grasp: consider the name for the item at the top of a pencil that removes the marks the pencil lead (graphite) makes. In American English this item is called an "eraser." In British English it is called a "rubber." The first name identifies the object in question in terms of its function, the second in terms of its material composition, or more broadly its structure. Most nouns of most languages identify objects in terms for function (e.g., *chair*). By contrast, physical science often identifies objects in terms of structure (*oxygen* is the element whose atoms have eight electrons and eight protons, with an atomic weight of sixteen, except for isotopes).

The next step in the argument for the autonomy of higher-level science is to consider the apparently silly question, can we “reduce” the type, the kind, the concept of *chair* to purely physical concepts, that is, define it in terms of the physical structure that all chairs share? But chairs do not share much or perhaps even any physical structure: they don’t need four or three legs or even any legs (consider a solid throne). They don’t need backs or sides or even seats of a given size or shape. Chairs can be made of plastic, metal, wood, ice, plutonium, et cetera. Chairs do not have to support any particular weight or be any particular size (consider the chairs in a doll’s house). To be a chair can’t be reduced to any set of facts about the structure of chairs. Yet, no one would deny that chairs are wholly and completely physical things. No one would suppose for a moment that, just because we cannot define *chair* in terms drawn from physical science, chairs are nonphysical. No one has ever gone in for dualism about chairs, even though we cannot exhaustively break down the concept of “chair-ness” into more basic physical properties.

It will be convenient to have a few technical terms for the relationship between chairs and their physical constituents: *supervenience* and *multiple realizability*. It seems safe to assume that things such as chairs are “nothing but” physical objects, even though the concept of chair cannot be defined or described completely in terms of some set of physical properties shared by all chairs. There is a sense in which chairs are higher-level entities that supervene on lower-level ones in this specific sense: (a) given any particular higher-level entity—say, a particular chair—it will have a particular composition, a certain number of legs and arms, a seat and back perhaps, arranged in a certain way; (b) anything else whatsoever that has exactly the same material, physical composition, parts, et cetera, must also be a chair with exactly the same functions; (c) there will of course be other ways that chairs can be composed, perhaps even an indefinite number of other ways of being a chair. So being a chair is *multiply realized*. Therefore, we cannot ever complete a list of physical ways of being a chair that would be required to reduce the concept or kind of thing—chair—to more basic, purely physical kinds of things. But the crucial thing is that each individual chair is still nothing more than, nothing above and beyond, nothing greater than the sum of the physical parts that compose it.

Now this fact of supervenience is going to be true of almost anything that is defined in terms of its function, instead of its structure. It’s not just chairs and tables that supervene on an endless list of different ways matter can be put together to make them. Human social institutions are almost all defined in terms of their function, and not in terms of their composition. This will be obvious for, say, families and juries, private property or voting, money or corporations, markets or armies. In fact it's almost impossible to find a term of interest in the social sciences that is not defined in terms of its causes or (much more usually) its effects, especially ones that meet the needs of some individuals or groups (or something that itself supervenes on groups). If things in the domain of any one of the social sciences are defined in terms of their function, then presumably they will supervene on the actions and behavior of individuals and groups whose behavior is in the domain of some other social science. But that means that, although institutions are composed of nothing but the actions of the people who participate in them, they won't be reducible to them! Just as chair is not reducible to a list of physical components, being a jury or a market or marriage contract is not reducible to some set of people's particular behaviors or beliefs or desires, or any combination of them. Each and every higher-level social fact may be composed of lower-level facts, even individual psychological facts. But there are so many different ways that the same higher-level fact can be realized by, instanced by, composed of lower-level facts, that the higher one can't be reduced to the lower-level one. It's multiply realized by a vast "disjunction" of different ways people's behaviors can make it up.

And of course it's not just that economic or political processes, institutions, and facts, can't be reduced to sociological ones, and sociological ones can't be reduced to psychological ones, owing to multiple realizability. Psychological kinds of states like beliefs and desires are not reducible to neural states of the brain, even though each and every belief and each and every particular desire that occurs to a person is nothing but processes occurring in their brains.

Now what all this means is that each of the social sciences will be autonomous and irreducible to some lower-level science, say, psychology, just as Durkheim held. And this won't merely be owing to ignorance or incompleteness of the higher-level science or the lower-level one. It will be due to the supervenience of the concepts, types, kinds in the domain of the higher-level science to the concept, types, kinds in the domain of the lower-level one.

Those social scientists who, following Durkheim, argue for the autonomy of their discipline from any more basic social or behavioral science, or for the irreducibility of social science to natural science more generally, have a very powerful argument from multiple realizability of functional concepts. But the fact that all the social sciences operate with functional and not structural or compositional concepts and terms, raises at least as many problems for the autonomy of social science from natural science, especially, from biology, as it appears to solve. It is to these problems we turn in the next chapter.

Introduction to the Literature

M. Martin and L. C. McIntyre, *Readings in the Philosophy of Social Science*, incorporates many of the most important canonical and contemporary papers on both the holism/individualism dispute and the debate on the nature and role of functionalism in social science. Durkheim's original argument is excerpted, along with classical papers by J. N. Watkins and Steven Lukes. Two more recent papers reprinted in Martin and McIntyre, "Reduction, Explanation and Individualism," by H. Kincaid, and "Social Science and the Mental," by Alan Nelson, are sophisticated discussions of reductionism and the autonomy of sociology from the psychological. J. Fodor, "Special Sciences (or: The Disunity of Science as a Working Hypothesis)," defends the autonomy of disciplines not reducible to more fundamental ones. This paper is reprinted both in Martin and McIntyre and in N. Block, ed., *Readings in the Philosophy of Psychology*, which contains several other papers relevant to the autonomy of psychology from neuroscience.

CHAPTER TEN

Functionalism as a Research Program

Holism and antireductionism are two theses about the social sciences that go closely together with another research strategy, *functionalism*. As we shall see, all three face the problem we have repeatedly identified in this book, of how to account for the "spontaneous order" of social institutions, the fact that they meet social and individual needs that no individual or group of individuals can meet even if they noticed the need, intended to meet it, and took steps to do so.

FUNCTIONAL ANALYSIS IN SOCIAL SCIENCE

As noted in the last chapter, methodological holism goes along with another "ism"—functionalism. Moreover, functionalism as an analytical strategy and an explanatory one was also first advocated, explained, and exploited by Durkheim. And we may easily adapt Durkheim's arguments for functionalism to support the methodological holism sketched in Chapter 9 as a substitute for his more polemical arguments.

Functionalism as an explanatory strategy is fairly obvious and common, both in ordinary life and in biology. We often explain something's character or even its very existence by citing the function it serves. The function something serves is one or more of its effects, or the effects of its presence and behavior. The presence or operation of something has indefinitely many effects, but only a few of them are among its functions. Thus, among the effects of my pressing the brake pedal on my car is to wear out a bit of the brake pad, to dissipate the energy of motion into the air as heat, and to cause tread to be burned off the tire. But none of these effects is its function—

slowing the car down. This function—slowing the car down—explains why I pressed the brake pedal, of course.

For a considerable period in twentieth-century social science, the methodology of seeking functions for features of human affairs was called *structural functionalism*. Among its leading advocates were European anthropologists such as Maus, Malinowski, and Lévi-Strauss, figures who influenced the tradition of philosophers labeled *philosophical anthropology* in Chapter 8. The label has become slightly misleading in light of the contrast discussed in the last chapter between identifying things by reference to their functions—their beneficial effects—and by reference to their structure, composition, component parts, et cetera. However, in social science the label *structural functionalism* is appropriate because it reflects the close connection between functionalism as a method and holism as a hypothesis in social science. *Structuralism* labels the thesis that there are features—structures—of society that are necessary for its persistence but not intentionally constructed or even noticed by its participants, and that therefore these structures can't be reduced to or explained in terms of the behavior or the thoughts of individuals and groups of them. Structuralism is a holist thesis. It mandates a search for the functions of the structures it identifies but excludes the possibility that these functions are intended, designed, or even recognized by the participants in a society. In the remainder of this chapter we will drop the qualifier *structural* and use the label *functionalism* with the understanding of its structuralist suggestions.

To understand the strategy of functionalism, consider some examples. "What's that rock doing in front of the door? It's a doorstop." "Why does the heart beat? In order to circulate the blood." As we have seen (Chapter 5, "Causation and Purpose"), both of these explanations are problematical: the ordinary one because it is intentional, and the biological one because it isn't! That is, explaining something's purposes in terms of our desires and beliefs introduces all the problems of intentionality—how do our beliefs and desires represent things, how do they have content? Explaining something in terms of purposes that no person has seems to require the desires of God, or some other intelligent agent that modern science would rather not have to invoke.

Problematical or not, functional explanations are as indispensible in social science as they are in biology. Consider the question of why the market price system is almost universal, even in societies where governments attempt to suppress them. The answer to this question is that the price system is universal in human societies because it fills a function: coordinating production capacities and consumption demands that no individual or group, even armed with a most powerful supercomputer, could achieve.

As we will see, the philosophical puzzle is not deciding whether functional explanations are legitimate. They are too widespread in social science for any philosophical argument to cast doubt on them. The puzzle is giving an analysis of how it is that something could have arisen and persisted in human society that fulfils a function so beautifully but that no one could have designed or could maintain. The puzzle is all the more pressing for almost every social institution fulfills such functions even though few were ever designed by people "on purpose" to fill their function or any function for that matter. This problem is the one Chapter 6 labeled the problem of spontaneous order.

FUNCTIONAL INDIVIDUATION AND SUPERVENIENCE

Along with explaining why things happen, appeals to functions have another role in the social sciences that is equally indispensible. They are used for "individuating" and classifying things, for identifying units or wholes, for tying together disparate and apparently unconnected things into a larger category. If we are classifying things in terms of their functions, it may be easier to uncover interesting generalizations about them. Often these are generalizations that would have escaped our attention were we to classify things by shape or color or size or composition. Recall our discussion of chairs in the last chapter. Or consider another example, the functional concept *clock*, meaning a system for telling time. Now consider the incredibly diverse set of physical objects that have been used to fulfill this function. There are, first of all, the many different kinds of watch mechanisms—escapement wheels, tuning forks, quartz mechanisms, microprocessors, et cetera. There are atomic clocks, cesium clocks, pendulum clocks, the human pulse, sundials, water clocks, hourglasses, marked wax candles, the sun, leaves that change colors with the season, tree rings, and so on. What do all these things have in common that makes them clocks? Certainly no simple common physical mechanism (except perhaps at the ultimate level of quantum mechanical description). What enables us to identify them all as clocks is their function, that is, the uses they can be put to. And we would be unable either to calibrate one clock mechanism against another or to improve on the accuracy of any one of them without the general functional category of clock. "Clock" is the only category that enables us to bring this physically heterogeneous collection together in one theory that explains their common behavior.

The functional concept of clock permits us to do something else: given a collection of "junk" on a workbench, just our learning that the objects go together to make up a clock will help us figure out how to put them together, what each of them does, what sort of a thing they compose, and so forth. In contrast, if we know little about the physical makeup, or construction, of things, the best way to begin to learn is to see what the different things do, what function the parts go together to perform, if they have a function. For this reason, knowing functions enables us to identify wholes made up of components and to tell which components go together to make which wholes.

Now, macrosocial science requires this sort of functional analysis just because it claims to be autonomous from psychology and the rest of science. The claim of autonomy is the claim that knowing about the behavior of individual people can't tell us much of anything about the social facts, because psychological theory is no help in developing sociological theory.

Why not? The reason was given in the previous chapter: the kinds that sociology individuates, that is, what it distinguishes as needing explanation or providing it—marriage rules, or social classes, or religions, or community networks, et cetera—are all multiply realized by the psychological facts about people that they supervene upon. Recall from the last chapter that multiple realizability is a feature common to things like chairs or clocks that share in common a function that can be and is usually discharged by any one of a large number of different structures or mechanisms. Just as there are many ways something can be a clock, there are many different packages of different individuals' psychological attitudes, beliefs, wants, hopes, fears, preferences, and the behaviors they bring about, which can go together to constitute a social class, or a religious ritual, or a cricket match. Even though these higher-level, functionally individuated, sociological categories supervene on lower-level psychological categories, they cannot be reduced to them, defined in terms of them, or explained by appeal to psychological laws or regularities, owing to their multiple realizability. The indispensability of functionally analyzed kinds, categories, and concepts in sociology is at the same time an argument for the autonomy of sociology, and of every social science that invokes functional analysis, from psychology, and for that matter from the natural sciences.

In fact, this argument suggests that the only route to understanding macrosocial processes is through functional analysis. We have to ignore the problem of "composition" or "structure," the question of what social facts are composed of, because knowing this information won't help us identify social facts or discover any generalizations about them. That leaves only the study of how they work, what they do; in short, their function as a source of sociological theory. And we can neither identify social facts nor discover the regularities that systematize them without assuming that the facts have functions.

THE INEVITABILITY OF FUNCTIONAL ANALYSIS IN SOCIAL SCIENCE

Almost all significant features of human affairs—historical actions, events, processes, norms, organizations, institutions, et cetera—have functions, that is, adaptations, or they are the direct results of such adaptations. The idea that almost everything of interest to social scientists has functions may seem dubious at first blush. How could almost everything in human affairs be an adaptation? That sounds like an idea worthy of Pollyanna or Voltaire's Dr. Pangloss, who thought that the bridge of the nose was there to rest eyeglasses on. Even in biology, not everything turns out to be an adaptation. Much of evolution is a matter of drift—the play of chance on small and sometimes even large populations that leads to changes in the distribution of adaptations, and even to the persistence of nonadaptive and maladaptive traits. Moreover, important biological traits are just the result of physical constraints—gravity, ambient temperature, the seasons, and the length of the day. Surely all the same must be said of the course of human affairs. Indeed, it's fair to suppose that there may well be a great role for the drift of random forces and physical constraints in human affairs, just as there is in biology.

Defending the claim that most features of human affairs have functions relies a great deal on the qualification *significant*. There will be many features of human affairs that are the result of drift, and yet few social scientists will accept the suggestion that what particularly interests them about human affairs is the result of random drift alone, or even mainly. Similarly, social scientists will recognize constraints of many kinds as forcing subsequent features of human affairs to adapt to them. But few social scientists accord such constraints with the fixed character that constraints—especially physical ones—have in biological evolution. This is much of the point of the widely shared view that social facts are constructed. In fact, the most revolutionary social changes break down the oldest, firmest, and most pervasive constraints, as a result of processes of variation and selection. The real issue is whether such variation is blind and the resultant selection natural.

Reflection on human affairs does suggest that, even more than in biology, significant features of social life have functions for some one, or some group, or some practice. Some social practices, norms, institutions have been constructed by individuals and groups to cope with an environment that has mostly come to consist of other individuals and groups and their practices. Social life is nothing but groups and their practices competing and cooperating with one another. Some of the functions these practices have are ones people think they designed—institutions like the US Constitution. But mostly the adaptations emerged from history without anyone intending, designing, or even recognizing them. This is especially true of the most important ones: think of feudalism, or the Roman Catholic Church, two institutions that were around for a long time. Then there are short-lived adaptations. Then there are the features of human life that no one designed, that didn't emerge unintentionally from actions and events people did "design" or intend, but that are best thought of as symbionts, or parasites, or sometimes combinations of both, living on human life, and changing it for the better or for the worse, but always adapting the way they function to ensure their own survival.

It may be difficult to think of tobacco smoking or heroin addiction as having social functions, because they are harmful. But they must still be understood in terms of their functions, as we can easily show. These practices and many other persistent ones are harmful to humans, but they are practices with features that ensure their persistence and spread through human history, at least until their environments change and their effects start to be selected against.

Chinese foot binding is a nice example of how this works. Foot binding persisted for about a 1,000 years in China. It got started because women with bound feet were more attractive as wives. Bound feet functioned in part as a signal of wealth, since only rich families could afford the luxury of preventing daughters from working. Another function of bound feet was to make it easier to keep track of girls, and ensure their virginity, et cetera. So, at first, when the practice arose, foot-bound girls had more suitors. Pretty soon every family that could afford it was binding daughters' feet to ensure they'd get married. Result: when every girl's feet were bound, foot binding no longer provided an advantage in the marriage market, and all foot-bound girls were worse off because they couldn't walk, suffered other health effects, et cetera. Foot binding starts out having function for some families, or even for some girls. By the time it becomes really widespread and fixed, it has lost its original function (of giving a few girls an advantage in the marriage market) and acquired another one (making a girl marriageable at all), even though it is actually harmful to every foot-bound girl's health and welfare. But once everyone was doing it, no one could get off the foot-binding merry-go-round. Anyone who stopped binding daughters' feet condemned them to spinsterhood. That is why foot binding persisted despite its harmful effects. For whom or for what did its features have functions? For itself, for the practice, norm, institution of foot binding! The practice persisted, like any parasite, because some of its effects had the function of exploiting the "weaknesses" of humans and their institutions—marriage, the desire for virgin brides and large dowries, the desire to control women before and after marriage.

Once we widen our focus, the claim that almost everything of interest to social scientists in human affairs has functions becomes far less Panglossian.

But can it be correct? One reason to suppose it must be is the fact that almost all the vocabulary and taxonomy of common sense and the human sciences are themselves thoroughly functional. As a consequence it would be difficult for common sense and social science to even notice or describe anything except in terms that attributed effects to it that are beneficial for someone or something! Moreover, the predictive and ameliorative goals of the human sciences impose upon all of them research programs that assume that most of the significant features of human affairs are adaptations for some individuals and groups, and maladaptations for other groups. Though each of the social sciences may be neutral on the functions of the actions, events, norms, practices, and institutions in the domains of the other social and behavioral sciences, it will not be agnostic about those within its domain. This will be true at least so long as it has ameliorative ambitions for social processes in its domain. The remodeling and redesign of political, legal, economic, social, or cultural institutions, rules, norms, and practices would be impossible if these items did not to varying degrees have functions for individuals, and groups of various sizes and compositions, or for themselves as parasites on human wants, needs, and interests.

Since the same function can be fulfilled many different ways, and many different variants on the same norms, practices, institutions, et cetera, can have the same function, discovering the function of some feature of human affairs may tell us little or nothing about the psychological processes and individual actions on which it supervenes. The multiple realizability of a single type of social fact, by a vast disjunction of different individual behaviors combines with the functional character of most of these facts to provide a strong foundation for methodological holism.

Methodological holism thus begins with functionalism and adds to it the reasonable hypothesis that what we have learned about individual behavior provides little direct insight about the functions of macrosocial phenomena. Once we have discovered systematic regularities, if any, about social facts, a psychological theory may be called upon to help explain how particular cases of the regularities play out from instance to instance. But psychology will not help us identify the basic units and kinds of social facts that are regulated by social forces. Nor will the concepts, kinds and descriptions of ordinary language give us the right taxonomy for sociology. The divisions that it identifies reflect functions, but not necessarily the ones science seeks, or even the real functions, and certainly not the basic functions of social institutions.

Thus, for example, the jury system is identified in ordinary life and the law as the institution (in nations employing the English common law) that has the function of determining matters of fact in legal proceedings. But this may be quite a superficial functional analysis, one that obscures some other, deeper functional role or disguises the fact that the jury system shares important functional properties with other institutions in countries without juries at all. Both of these possibilities are important because identifying "deeper" functions or wider functional categories is essential to uncovering interesting sociological generalizations. Thus, some Marxian sociological analysis may hold that the real function of the jury system is *ideological*—to encourage public acceptance of decisions made secretly elsewhere on the basis of class interests instead of on the basis of real guilt or innocence. The institution of the jury system is thus to be explained in terms of its "real function," that of sustaining the ruling classes. Alternatively, if the jury system's role is described widely as that of "peaceful conflict resolution," then it will be classified together with other social institutions having the same role. Subsequently, the sociological theorist will attempt to frame generalizations about conflict-resolving institutions. These will be generalizations that can be tested by further examination of the jury system and other institutions with the same function.

The difference between apparent and real functional roles is often described in terms of the distinction between *latent* and *manifest* functions. The manifest functions of a social institution are those that it was, as it were, intentionally designed to accomplish and that it is recognized by its participants as accomplishing. Latent functions are those it serves unwittingly, without the recognition of its participants. Such unnoticed functions are often held to be more important and more systematically significant than the manifest functions of the institution.

For instance, the manifest function of marriage is to legalize domestic and sexual relations and regularize the duties and rights associated with them. According to Durkheim, however, marriage has other latent functions. It is one of the many institutions that protect the members of society from suicide. Its latent function is that of maintaining the optimal degree of social integration. In that respect, it is to be grouped with other social institutions that may seem quite different from it: the jury system, for example, or the institutions of the Catholic parish.

Identifying things by their functional role may also enable us to recognize the artificiality of distinctions between social institutions. Sometimes these boundaries prevent us from recognizing generalizations that may explain them. For example, the functions of the police as the agency of law enforcement, and of the courts as the agency of factual determinations in legal questions, may seem quite distinct. Yet the Marxian sociologist who views the jury system as an institution for enforcing class interests may bring the jury system together with the police into one category of institutions with a single latent function. That sociologist may argue that the institution operates effectively by making it appear as though both its components have distinct identities and separate functions.

Individuating social institutions by function and framing explanatory theories about them are intimately associated. We would be unable to discover any generalizations about functionally identified social institutions unless we first identified them in terms of their functions. But just to identify something as an instance of a functional category is to advance a generalization about it. To identify marriage as a socially integrating institution involves asserting that marriage encourages social integration. That is a generalization as well as a classification. It enables us to lump marriage together with other such institutions and then to see whether we can frame hypotheses about them. For instance, we may consider the generalization that certain institutions reduce the individual's probability of suicide. In fact, Durkheim's claim that there are three different types of suicides—egoistic, altruistic, and anomic—is based on a prior classification of social institutions into functional types that protect against too much social egoism, too much social altruism, or social anomie. The classification of suicides was based on an examination of the consequences of three different ways that institutions can break down, failing to fulfill their function of maintaining an optimal degree of social integration. So functionalism is both an analytical strategy for identifying socially significant institutions and an explanatory strategy that accounts for institutions' characteristics by appealing to their effects for society as a whole.

As noted in the last chapter, if holism is correct, either as a doctrine about the independent existence of social facts or as a methodologically reasonable practice, then functional analysis and functional explanations are obviously appealing. In fact, they are more than appealing. They are indispensable. That is why holists are functionalists. However, functionalism is a method with some serious potential problems. These are problems that methodological individualism has seized upon in its counterarguments to both ontological and methodological holism.

THE TROUBLE WITH FUNCTIONALISM

According to the methodological individualist, the commitment to functionalism represents everything that is wrong with holistic social science. Functionalism is held to be complacent at best, immoral at worst, and sterile when it isn't untestable altogether. Besides, individualists charge, it rests on a false view about the nature of society, the view that society is some sort of organism, as opposed to a collection of "atomic" individuals. Accordingly, individualists recommend we turn our backs on the search for latent functions and search instead for explanations of social phenomena that appeal only to the behavior of individuals. When we fail in the employment of this strategy, we should blame our own lack of scientific ingenuity and not twist the facts to explain our failure.

According to the individualist, functionalism works as a method in biology because the subject matter of biology is organisms—and their organs, tissues, cells, and so on—which have indisputable functions with regard to the survival and well-being of the organism or its environment. Holism and functionalism are tenable only on the assumption that society is some sort of superindividual organism, made up of institutions and individuals acting as its organs, tissues, and cells. But, the individualist insists, society is not an organism, and there are scientific and moral dangers in even the metaphorical treatment of it as an organism. Therefore, these "isms"—holism and functionalism—encourage cognitively and morally dangerous social science.

First consider the individualist's morality charges: at its worst, holism is an accomplice of totalitarianism of the right and the left. Holism and functionalism, by according social institutions a life of their own and attributing to them functions with respect to the needs of the society—as opposed to the needs of the individuals who compose it—threaten the priority of personal liberty and individual human rights. For example, we hold that the jury system has the function of ensuring the rights of the accused. The suggestion that it has some other deeper, latent function undermines the priority of the protection of rights as its real function and encourages us to view this institution as serving some other needs, ones with social priority over the protection of individual rights. If the real function of elections is, as Marx put it, for the proletariat regularly to pretend to decide which among the capitalist classes will exploit it, then someone who adopts that Marxian theory is unlikely to respect the process or outcome of "free elections."

It is regrettably true that totalitarian political philosophies, from Plato to Marx, have subordinated individual rights and advantages to the needs and well-being of society. They have justified the subordination of individual rights on a holistic theory of social organization, one that makes society as a whole into an agent with rights, claims, and interests. But it is also clear, to many holists at least, that this misuse of a version of their methodological prescription is no reason to condemn all uses of it. For, as some of them argue, their theories are value free, are neutral on moral and political applications, and certainly embody no prescriptions about how society should be organized.

Nevertheless, the individualist replies, holists and functionalists must be inclined by their doctrine to be complacent about social arrangements. At the very least, functionalism is an unintended bulwark against social change. The identification of social institutions in terms of their function carries with it the implicit suggestion that they fulfill a need of society, a need, on the latent function theory, that we may not have recognized. Therefore, one must be leery of replacing institutions or changing them considerably. For the change may adversely affect society's ability to meet its needs. Conservatives often point to the unintended consequences of social change, which may overwhelm the foreseen ones. Functionalism is grist for their mill, since it holds that beyond the things institutions do directly for individuals, they do things for the society that social planners often fail to take account of when they set out to make "improvements."

This is a charge that may have more substance than the complaint that holism is akin to totalitarianism. For functionalists are likely to seek support for their method in an evolutionary approach to society, one that identifies institutions as adaptations selected by nature. To call something an adaptation certainly seems a way of commending it.

That brings us to the methodological objections the individualist offers against holism. Treating society as an organism, even metaphorically, and taking latent functions seriously force the holist to make difficult choices. The holist must either opt for Durkheim's *âme collective*—the group mind—to explain how society arranges institutions to meet its needs, or embrace a Darwinian evolutionary view, according to which all long-lasting social institutions arose through variation and selection for their beneficial functions. That is an alternative to which we shall return in the discussion of sociobiology in the next two chapters.

In some ways, functionalism is a natural development of the strategy of finding meaning in human affairs. One reason it is so appealing is its similarity to folk psychology's approach to explaining individual behavior in terms of purposes that give actions their meaning for us. By finding latent functions in our actions or institutions that we do not recognize, functionalism provides the resources for other theories that seek deeper meanings. Thus, for example, it can help defend the search for meanings against the charge of banality that anthropologists face.

Functionalism is in some respects a far more appealing approach to deeper social meanings than psychoanalytic theory, and it is a natural way to interpret Marxian theories. Freud's account of deep meanings is only one of a wide range of functional theories from psychoanalysis that can be applied to social institutions generally. One can still search for deep meanings even if one repudiates orthodox Freudian psychoanalytical approaches to them. Instead of investing unconscious psychological states with unrecognized Freudian purposes to explain action, one invests social institutions with such purposes and then shows how the institutions constrain, overwhelm, or inform individual action with a deeper meaning, derived from their institutional function. We have seen this strategy at work in some of the philosophical anthropologies discussed in Chapter 8.

The attractions of functionalism cum holism as a way of interpreting Marx's theory are evident. Society is to be viewed as a system composed of classes competing for supremacy. The institutions of society are analyzed in terms of their functions in fulfilling the needs of the competing classes. The Marxian critique of ideology identifies the meaning of aspects of the ideological superstructure in terms of the interests, not of individuals, but of social classes, that these aspects serve. The whole society is itself viewed as a superorganism composed of these classes, one that changes over time in ways that are self-perpetuating. Of course, not every aspect of Marxian doctrine can be easily accommodated to this approach. In particular, it is difficult to reconcile functionalism with the reflexive character of the theory in which critical theorists set so much stock. But it is not impossible. Nevertheless, we can leave this matter to Marxian scholarship, for there are both holist and individualist strains in Marx's writings. Functionalism provides Marxism's successor theories of gender, race, and class conflict with their chief methodological tools. Features of society not recognized by its participants as racist or (hetero) sexist or exploitative can be identified and classified together and given their real meanings on the basis of their roles in maintaining capitalist-dominated, racially privileged, or patriarchal institutions.

The question the individualist raises is, For whom or what are these meanings significant? Not for individuals, for the meanings of institutions are not to be found in individuals' subconsciousnesses or in their immediately self-identified interests. If functions provide meanings that explain institutions, then we need an intentional agent in which to "locate" these meanings—unless, of course, talk of meanings is metaphorical, or figurative, in these contexts. For Durkheim in *Suicide*, at least, there was such a social mind or spirit, and his arguments for it are far from derisory. Nevertheless, few have followed his advocacy of collective consciousness.

HOW CAN WE JUSTIFY FUNCTIONALISM IN SOCIAL SCIENCE?

We therefore need another rationale for the attribution of functions, one that doesn't require a mind to which they are meaningful. Recall that functional attributions and explanations are teleological: a system's functions are some of its many effects. They are those among its effects that meet some need, accord some benefit, confer some advantage, either to the system or to some larger system that contains that system with the function in question. So functional explanation is explanation of causes by their effects. As our reflection on the problem of teleology (Chapter 5, "Causation and Purpose") shows, this presents functional explanation with a serious if not fatal problem.

If a functional theory is to be a causal one, then it cannot allow later effects to explain earlier causes. Causes precede their effects; they never follow them! Functions are later effects, not explanatory prior causes. For example, Durkheim tells us that one of marriage's functions is the reduction of alienation and anomie. The evidence for this claim is the reduction in the incidence of suicide among married persons. But results, such as the reduction of anomie, and the consequent reduction of suicide rates, cannot cause their antecedents—marriage, nor can they causally explain them. And, of course, some married persons commit suicide; thus, in these cases, marriage did not have the usual effect, yet that does not detract from Durkheim's functional analysis of the role, significance, meaning of marriage in society.

This problem of how functions—later effects—can explain the presence or persistence of traits is one that biology faces too. The solution in biology, as we have seen, is an appeal to the mechanism of evolution by natural selection. The function of the heartbeat—to circulate the blood—means that, over the course of evolution, random variations in heart configurations that fostered circulation were selected for because of their contribution to the fitness of the animals with those variations. The selection of successive variations produced hearts of the current adapted design. Thus, a heart that circulates the blood is an adaptation. Functional claims in biology turn out to be only apparently about immediate effects and really about ultimate prior causes in the long evolutionary past.

Thus, in biology, claims of the form, The function of *x* is to *F,* are invariably backed up by an *etiology*—a history of heritable blind variations that the local environment filters among for fitness, or adaptedness, for having more or better *F*-capabilities. This persistent cycle of variation and filtering produces a succession of increasingly adapted *x*-like structures that eventuate into *x*'s that do *F.* So in biology, to attribute a function to a biological trait is often to commit oneself to a Darwinian process as the etiology of how the trait came about.

Sometimes functional analysis in biology and especially in physiology does not explicitly presuppose Darwinian etiologies of blind variation among heritable traits and successive rounds of environmental filtration that sculpts them. Sometimes to attribute a function to something is simply to assert that what it does contributes to the capacity of some larger structure containing it to behave in a certain way. Thus to say that the function of the iris is to modulate the amount of light falling on the retina has no explicit implication that it was selected for exactly this and only this effect, or that it was a single trait or feature of the eye selected for at all.

When "function" is used in this sense in biology, it seems free from any explicit suggestion of why the trait or its function is advantageous for any biological entity at all. However, when function is used in social science, there is always this suggestion of benefit conferred or adaptation. So this "causal role" account of how the concept of function sometimes works in biology has little relevance to the present problem. We still need to know by what right the holist, or any social scientists for that matter, attributes a function, especially a latent one, to a social institution.

Well, why can't sociological functionalism take a page out of biology's book? Why can't it help itself to a Darwinian theory of the natural selection of societies and their traits—their institutions. Such a theory might go something like this: blind variation produces a variety of social institutions with diverse effects for the future of societies. Among these institutions, some were adaptive for the societies in which they arose, some were maladaptive, and others were neutral. Societies with adaptive institutions flourished, those with maladaptive ones extinguished themselves, and those with adaptively neutral institutions were overwhelmed in competition with societies with the better adapted ones. By a succession of refinements through the mechanism of variation and selection, there arose societies with the institutions we recognize today, since all these institutions are adaptive for the societies in which they occur. And this entirely causal account underwrites the functional analyses and explanations the holist requires.

In the first three-quarters of the twentieth century, hardly any social scientists were tempted to offer such a Darwinian theory for the existence and character of social institutions, social facts about these institutions, and their relationship to one another. At the same time it must be admitted that there are no other credible competing explanations for the emergence and persistence of holistic social facts about institutions with functions. Very few social scientists, including most holists, ever took the Darwinian theory seriously. However, few holist social scientists ever seem to express disquiet about the absence of an explanation even for how social facts about holistic social institutions with functions are possible.

One reason few social scientists were prepared to take Darwinian theory seriously for social science was the widespread belief that, as a theory about genetically fixed, hardwired traits, it had no relevance to human affairs. Almost nothing of interest in human affairs seems to be hardwired and genetically encoded; human societies and their components are so different from one another, while human genetic inheritance is so similar, that it's obvious the latter could not explain the former. Culture and civilization, it was confidently held, is a matter of nurture and not nature. Much of it is the product of cognitive processes that are themselves acquired by learning, that is, are transmitted by linguistic and other culturally dependent processes. Thus, the very idea that a Darwinian theory of natural selection might be relevant to human affairs was considered laughable when it was not stigmatized as dangerous. It was viewed as dangerous, since theories of human cultural differences as genetically hardwired were convenient beliefs for racists, sexists, and xenophobes.

There was a further and much more powerful reason why Darwinian theory was long deemed of no relevance to human affairs and therefore unable to provide any support for a holistic theory of the function of social institutions. The most important feature of human society is the fact that people cooperate, that they behave in accordance with moral norms that prevent people from acting selfishly, egotistically, and without regard for other fellow creatures. Admittedly, these norms are sometimes enforced only within groups and not across them, and there are always a handful of individuals who flout these rules. But no society could function without substantial cooperation among its members; no social institutions could long exist without all participants sharing the burdens of its maintenance; and there would be no social facts about institutions to explain without human cooperation. But Darwin's theory apparently could not explain cooperation among humans. Indeed, it would lead us to expect that there is no cooperation among them, as we shall see. The evident falsity of this seeming implication long made Darwinian processes completely unavailable for any account of function in social science.

The issues raised by arguments against the relevance of Darwin's theory of natural selection to the social sciences became so important in the last third of the twentieth century that outlining them will require most of the next two chapters. Meanwhile, the main conclusions that need to be emphasized about functionalism in the social sciences are these: First, almost everything of interest in human affairs has functions and indeed most things of interest in social science are identified by mentioning their functions. Second, the fact that something has a function requires explanation. In the case of artifacts with functions we designed them to have, the explanation is obvious. But few social institutions and fewer of their functions are the results of conscious intentional design by anyone. Unless the fact that most everything in human affairs has a function is taken to reflect God's benevolence, there has to be some explanation of this fact about human affairs. Without a credible explanation, in fact, we should conclude that functionalism is mistaken. We would have to conclude that institutions don't arise and persist because they have functions, and the belief that they do is a mistake or illusion.

Introduction to the Literature

Durkheim's views are expounded in his extremely important manifesto, *Rules of the Sociological Method*, and they are illustrated in *Suicide*. After Durkheim, the most prominent early advocates of functionalism in sociology are B. Malinowski, *A Scientific Theory of Culture*, and A. R. Radcliffe-Brown, *Methodology in Social Anthropology*, who coined the term "structural functional." In American sociology, the method is closely associated with the work of Talcott Parsons, *The Social System*. The latent/manifest function distinction is owed to R. K. Merton, *Social Theory and Social Structure*. A more contemporary and sophisticated version of functional theory is to be found in Jonathan Turner's *Theoretical Principles of Sociology. Vol. One: Macrodynamics*, and in Turner and Stets, *The Sociology of Emotions*.

Problems for functionalism in sociology are lucidly identified by C. Hempel's "Logic of Functional Analysis," in his *Aspects of Scientific Explanation*, and reprinted in Martin and McIntyre. These issues are identical with those surrounding teleology (see Chapter 5, "Causation and Purpose").

The most vigorous opponent of holism and advocate of methodological individualism has been K. Popper. See, especially, *The Poverty of Historicism* and his attack on the moral foundations of holism, *The Open Society and Its Enemies*, volumes 1 and 2. Popper's doubts about holism and functionalism extend even to evolutionary theory and to Freud as well as Marx.

Much of the debate about the propriety, logic, and foundations of functional explanations has been carried out in the philosophy of biology. For an introduction to this debate, see A. Rosenberg, *Structure of Biological Science*, and an anthology of significant papers, E. Sober, ed., *Conceptual Issues in Evolutionary Biology*. Some social scientists embrace a functional approach while explicitly abjuring any causal mechanism underlying it. See R. Needham, *Structure and Sentiment*. The role of individualism and the need for underlying mechanisms in a functional interpretation of Marxian theory has for some time been a lively philosophical issue. See Daniel Little, "Microfoundations of Marxism," G. A. Cohen, "Functional Explanation in Marxism," and Jon Elster, "Functional Explanation in Social Science," all in Martin and McIntyre.

CHAPTER ELEVEN

Sociobiology or the Standard Social Science Model?

Breakthroughs in evolutionary theory combined with long-standing problems—mainly facing holism and functionalism about social practices and institutions—led in the last thirty years of the twentieth century to the increasing influence of biology in the social sciences. The result was a long period of debate about the relative importance of "nature versus nurture" in human affairs. This chapter traces the debate's enduring impact on the philosophy of the social sciences.

SOCIOBIOLOGY OR THE STANDARD SOCIAL SCIENCE MODEL?

Chapter 10 made it clear that holism and functionalism are two approaches to the explanation of social phenomena that go hand in hand. The chapter also expounded the problem facing the combination of holism and functionalism: since almost everything of interest to the social scientist has a function, in the absence of divine or human designers it is difficult to see where these functions come from. Holism, of course, rules out the simple combination of human intentions and actions as the source of institutions' functions. That is the approach of the methodological individualist.

Holists, therefore, are often functionalists, but urgently require an explanation for how social structures arose that are so conveniently functional for either the preservation of society or the benefit of its members. They also need an account of the persistence of such benefits and their resistance to challenges to their survival. These interesting social facts cannot have been miraculous coincidences, happy accidents, especially not when as holists (and almost all social scientists) believe, human affairs is rife with well- (or ill-) functioning institutions, norms, practices, policies, artifacts, et cetera. Durkheim and those who followed him recognized this fact about societies. It led them to conceive of social wholes of various kinds as quasi-organisms, or even, in the view of some social scientists, as literal, real organisms, living things, whose "cells" were the individual people that compose them. As a metaphor this new "ism"—organicism—had its advantages, like all good metaphors in science. But it is obviously difficult to take it as more than a metaphor, unless we can apply biological theory to human affairs in such a way as to vindicate the literal claim that societies are organisms.

The irony is that when serious attempts began to be made to do this very thing—apply theory vindicated in biology to social science—the result was to sever the connections between holism and functionalism. What is more, it strongly undermined holism, while providing a completely unexpected justification for functionalism. And it prompted a revolution, a "paradigm shift" that quickly worked its way across all of the social sciences, with repercussions for each of them and for the philosophy of social science still very much in debate today.

To explore how this all happened we need to begin with the work of a surprising advocate of holism in human affairs: Charles Darwin.

DARWIN'S "HOLISM" AND ITS PROBLEMS

Charles Darwin didn't have to argue that biological organisms had traits with purposes, functions, and adaptations. This much was obvious and was a premise in the "argument from design" for the existence of God. Darwin, of course, discovered the real process that produces the appearance of design and produces the beautiful means/ends economy of parts to wholes and organisms to environment that previous thinkers had mistakenly attributed to God's benevolent design. The process Darwin discovered was natural selection.

The theory was so simple that Darwin's first great supporter, Thomas Huxley, reacted to his reading of *On the Origin of Species* by saying, "How stupid of me not to have thought of it!" This reaction is not surprising. In hindsight, the constraints on any possible explanation of the appearance of function or adaptation in nature are so narrow that only a theory like Darwin's can fulfill them.

As noted first in Chapter 5, in the absence of free-floating goals, purposes, or future ends, there must be some prior causal mechanism that brings about these functions. Darwin hit upon what looks like the only mechanism that could do this: blind variation and natural selection. Certainly, no one has produced a plausible and empirically supported alternative in the 150 years since the publication of *On the Origin of Species.* The theory has only three fairly modest tenets:

1. There is descent, in which later generations have hereditary traits more similar to their predecessors than to others.
2. There is always variation in every generation among these hereditary traits.
3. There are differences in the fitness to the environment of these hereditary traits.

Provided these three assumptions are realized by a domain of individuals, there will be what Darwin called "descent with modification," or evolution in the direction of persistently greater adaptation to the local environment. Where and when this process persists in a given environment for long enough, the result will be a very high degree of adaptedness of a thing and its parts to its environment. In other words, the parts will have been selected for their effects on fitness, and they will have become functions! Darwin had animals in mind, of course. But biologists include genes, animals, families, and populations, and, besides the obvious anatomical and physiological traits, they include behaviors, capacities, abilities, dispositions, and their extensions into the environment, like spider webs and beaver dams. And, of course, it was Darwin himself who first attempted to extend his theory to the social sciences.

From the time Darwin wrote *The Descent of Man* there have been repeated attempts to apply his theory of natural selection to explain social phenomena and human behavior. These initiatives have made little headway, for two main reasons. First of all, it was difficult to see how a theory about random variation and natural selection of genetically inherited traits could shed much light on the learned behavior of humans or their social and cultural consequences. Second, and perhaps even more important, the Darwinian mechanism operating over geological epochs works relentlessly to produce organisms designed for individual fitness maximization. But humans don't seem to be very good biological fitness maximizers, acting relentlessly to pass on the largest numbers of their genes in offspring. In fact humans look like counterexamples to Darwin's whole theory: 3.5 billion years of evolution relentlessly selecting for organisms that maximize their own fitness should have selected for look-out-for-number-one selfish egoists, who do anything each of them can to survive and reproduce in competition with every other organism, since all were the result of the same survival-of-the-fittest process.

But the overwhelmingly obvious fact about human cultures and social groups is that cultural norms and social institutions foster and enforce cooperation, sharing, altruism, promise keeping, respect for the interests of others, and a sense of justice and fairness in human dealings with one another that no purely fitness-maximizing creature could possibly honor, endorse, or obey. It looks like from the Darwinian point of view, human sociality, human society, human political and economic institutions should all prove to be impossible in the long run. For they all require cooperation, trust and promise keeping, unselfishness and other fitness-reducing actions, preferences, habits, and dispositions. Natural selection should condemn the people who act this way to long-term, indeed perhaps even short-term, extinction. Since it is evident that people just are not fitness maximizers, most social scientists concluded that when natural selection finally got around to producing Homo sapiens, it made a species smart enough to slip off the leash of the genes and transcend Darwinian constraints on evolution. Accordingly, these social scientists considered it safe to disregard Darwinian theory in the projects of the social and behavioral sciences. The trouble with this conclusion is that Darwin's theory offers the only scientifically acceptable explanation of how anything can have a function. Rejecting Darwinism about human affairs requires either giving up functional analysis or devising an alternative theory of how human institutions, et cetera, came to have them.

In *The Descent of Man* Darwin recognized the problem of reconciling the process of natural selection for fitness maximizing with the universality of cooperative institutions in all human societies. And he offered a holistic solution to it. The problems facing Darwin's solution reveal how difficult it is for Darwinian natural selection theory to accept methodological holism. But in the end it also enables Darwinian theories of human cultural evolution to reconcile fitness maximization and human cooperation, and provide a theory of how social institutions with functions can arise.

Darwin argued that the three-step process of natural selection, operating on generations of individuals and their traits, also operates on lineages of groups of people. Groups have traits—norms, institutions, practices, cultures—which are transmitted from generation to generation, and so are heritable. And these traits differ in their fitness—that is, in differing environments some are adaptations and others are maladaptations—for the groups that bear them. Differential fitness of inherited traits is all one needs for natural selection. Groups with cooperative members will be selected for. Why? When these groups competed with mega fauna and with one another for territory or prey, the winning groups would have to be the ones whose members worked together, cooperated with one another to fight off predators, kill prey, and eventually kill off other groups of people. Individual altruists may be selected against—heroes who throw themselves on grenades leave no offspring. But groups without people ready to sacrifice their interests for one another at least some of the time will lose out in the struggle for group survival. In the long run of competition between groups, those groups composed of people nice to one another would have been selected for. After long enough the only groups left would be those composed of people nice to each other. Whence, according to Darwin, cooperation, indeed altruism must triumph as fitness enhancing for groups, despite its immediate fitness costs for the individual members of groups.

Thus Darwin was committed to a version of methodological holism. Natural selection on his view operates at several levels independently: on the level of the individual organism, and on the level of groups of individual organisms. And it may be pushing individuals and groups in different directions, selecting for selfishness at the level of individual organisms, and selecting for cooperation at the level of groups of organisms. Since almost all human societies that have survived until today are composed of cooperative people, group selection must have won out over individual selection at least in the human case. Darwin's explanation of the emergence and persistence of cooperative institutions is holist at least in part because it assumes that individuals would not otherwise be selected for cooperating, so the group has to somehow make them—all of them—behave otherwise.

Most biologists eventually saw serious problems with Darwin's holistic theory of the evolution of cooperation. First of all, groups don't reproduce the way individuals do. They emerge and persist as members are born and die, but they rarely make copies as natural selection seems to require. Another thing we know about natural selection is that variation in traits is always the rule, never the exception. In fact, in every species and every subspecies, group, family, individual line of descent, and in every generation, there is always random variation of traits. In biology, these variations are often the result of genetic mutations. Whatever their source, it was Darwin's great achievement to realize that variation was universal, unavoidable, and the source of the changes, beneficial and harmful, that make a difference between which individuals survive to reproduce and which do not. Variation, however, will always undermine group cooperation, and this fact is fatal to Darwin's explanation of how cooperation emerged and persists.

Suppose group selection has operated long enough to make extinct every group composed of noncooperators, and left only groups composed of cooperators, nice guys, at least nice to their fellow group members. The trouble is, groups composed of nice individuals are always vulnerable to variation in the behavior of new members. A mutation can arise at any time among the offspring of nice people that makes one or more of them start to exploit the niceness of everyone else, to feather their own nests without being nice in return. Through their exploitation of the nice people around them, they will secure disproportionate resources. This improves their survival odds and increases the number of their offspring. After enough generations of such within-group competition, the original group of nice people has become a group of not very nice people on none of whom it ever pays to turn one's back. The result is that groups of cooperators can be expected by the process of natural selection alone to continually evolve into groups of noncooperators.

The problem of subversion of cooperative groups from within was one Darwin and other exponents of group selection were never able to solve. It left them, and Darwinian theory, with no cogent explanation of the universal fact about human cultures that they all manifest spontaneous and persistent cooperation. Since all the norms, practices, and institutions of every society are built on cooperation, it appeared that Darwinian theory had no bearing on, and was irrelevant to, human affairs and the social sciences.

Notice also that the problem Darwin's holistic explanation of cooperation faces reflects the methodological individualism of biology. Many biologists are leery of higher levels of organization that cannot be explained by processes at lower levels. They are, in effect, methodological individualists: the argument just sketched against Darwin's group selection theory is a methodological individualist one. It argues that selection at the level of individual group members determines the properties of the group, and not the other way around, as group selection requires.

Economists, who are particularly strong advocates of methodological individualism, may take some comfort from the fact that biology shares with economics this commitment to individualism. But what they did not notice was that the very same problem Darwin faced—explaining human cooperation—also seems to vex rational choice theory. The hypothesis that all human agents maximize their preferences, together with the fact that almost everyone prefers to consume more goods than fewer, to work less rather than more, and to benefit themselves more than to benefit others, should make the cooperative character of human society equally mysterious. The fact that natural selection for fitness maximizers should in fact result in individuals who are rational preference maximizers is not surprising. For what better way to assure the representation of one's genes in future generations than to secure with maximal efficiency the most resources with which to maintain health, power, and attractiveness to the opposite sex, and to support and nurture offspring in the competitive struggle for existence that nature establishes. The Darwinian theory of evolution by natural selection and the economist's theory of rational choice should really be in the very same boat when it comes to attractiveness as explanatory theories of society and human relations, and the boat should not, after all, be afloat.

The Darwinian biologist's problem of how groups can enforce cooperative traits when individuals are selected for rational preference-maximizing egoism is the same as the economist's problem of the origins and persistence of spontaneous order we first encountered in Chapter 6.

THE PRISONER'S DILEMMA TO THE RESCUE?

To see the problem that economic theory and evolutionary theory both face, consider the most well-known strategic interaction problem in game theory: the prisoner's dilemma (PD). Suppose you and I set out to rob a bank by night. However, we are caught with our safecracking tools even before we can break into the bank. We are separated and informed of our rights as criminal suspects and then offered the following "deals": If neither of us confesses, we shall be charged with possession of safecracking tools and imprisoned for two years each. If both confess to attempted bank robbery, a more serious crime, we will each receive a five-year sentence. If, however, only one confesses and the other remains silent, the confessor will receive a one-year sentence in return for his confession, and the other will receive a ten-year sentence for attempted bank robbery. The question each of us faces is whether to confess.

Only a little thought is required to see that this problem is easily solved. As a rational agent, I want to minimize my time in jail. So if I think you're going to confess, then to minimize my prison sentence, I had better confess, too. Otherwise, I'll end up with ten years and you'll get just one. But if I confess and you don't, then I'll get the one-year sentence. Now it begins to dawn on me that whatever you do, I had better confess. If you keep quiet and I confess, I'll get the shortest jail sentence possible. If you confess, then I'd be crazy not to confess as well, because otherwise I'd get the worst possible outcome, ten years. So I conclude that the only rational thing for me to do is to confess.

Now how about your reasoning process? Well, it's exactly the same as mine. If I confess, you'd be a fool to do otherwise, and if I don't, you'd still be a fool to do otherwise.

The result is we both confess and each gets five years in the slammer. Where's the dilemma? It's best seen in a diagram of the situation: the top of the box labels your choices: confess or don't confess. The left side labels mine: confess or don't confess. The numbers in the lower left of each square

are the number of years I'd serve under each combination of choices, and the numbers in the upper right of each square are the numbers you would serve. Each square is labeled in roman numerals for reference.

		YOU	
		Don't confess (cooperate with me)	Confess (defect from me)
ME	Don't confess (cooperate with you)	2 II 2	1 I 10
	Confess (defect from you)	10 IV 1	5 III 5

The prisoner's dilemma.

The rational strategy for you and the rational one for me lead us to square III, where both of us confess. These are called the "dominant" strategies in game theory because, as the reasoning shows, they are the most rational for each of us, no matter what the other person does. They dominate all other strategies. But now step back and consider the preference order in which each of us would place the four alternatives. My order is I–II–III–IV; in each successive square I get more years in jail. Your order is IV–II–III–I, for the same reason. Though we end up in square III, we would both prefer square II because we would prefer both of us getting two years to five years in jail. Yet rationality, maximizing our utility, leads us to a suboptimal outcome, one less desirable than another that is attainable. The dilemma is this: there is no way we can rationally get to square II, even though both of us rationally prefer it to square III. The reason is easy to see.

Suppose that before starting on the bank job, we both took oaths not to confess. If either of us believed that the other party would live up to the promise not to confess, confession would be even more tempting, for it would increase the chances of getting the lightest sentence by confessing. Suppose we backed up the promise by hiring a hit man to shoot whoever confesses and gets out of jail first. Then, of course, the rational thing is to make a further secret payoff to the hit man not to carry out his job, and then to confess anyway. In short, there seems no way for rational agents to secure a more preferred alternative. This then is the dilemma: given the payoff rankings, the agents, by trying to maximize utility, are prevented from cooperating to attain a utility-maximizing alternative.

Why should this toy model of strategic interaction pose a serious problem for rational choice theory, still less for an evolutionary account of the very possibility of social cooperation? Take the biological problem first. A prisoner's dilemma is any strategic interaction in which there are two choices for each agent, and the rankings of the payoffs are in the same order as that given above (I–II–III–IV versus IV–II–III–I), but the dominant strategy takes both players into box III. And if the payoffs are reproductive opportunities, the prospect of animals of all kinds finding themselves in a prisoner's dilemma is considerable. Consider the case of two scavenger birds who come upon a carcass. They could both fight to decide which will have the carcass to itself, during which time a third scavenger might steal it away; either one could defer to the other, which would reduce its fitness and enhance that of the other scavenger; or they could both start consuming the carcass. The trouble with this last prospect is that were either to consume unilaterally, the other would be able to attack it, perhaps fatally, thus enhancing its fitness by disposing of competition and securing the resource. Assuming that birds cannot negotiate an agreement to share the carcass, and that even if they could there would be no reason for either to trust the other, the scavengers face a prisoner's dilemma. Accordingly, they will not share the food, but both will warily stalk each other, neither eating nor fighting, until the carcass rots. Between two early hunter-gatherer human beings the same problem might emerge as well.

More seriously, we find ourselves in prisoner's dilemma situations repeatedly throughout life. Consider how often you have purchased a soft drink over the counter at a gas station just off the freeway or motorway in a region of the country you will never visit again. You have a five-dollar bill in your hand and want a drink; the salesperson at the counter has the drink in his hand and wants the bill. Your best strategy is to grab the drink, keep the five dollars, and drive off. His best strategy is to take your money and hang on to the drink. If you complain, he will simply deny you paid him. You won't call the police. You simply don't have time and it's not worth the trouble. Better to just go to the next convenience store on the road. Suppose you grab the bottle, pocket your bill, and drive off. Will the salesperson call the police? If he did, would they give chase? Could they identify your car? The answer to each of these questions is no. It's not worth their trouble. Knowing all this, neither of you does the rational thing; you thoughtlessly, irrationally cooperate, exchanging the five-dollar bill for the drink. We can multiply examples like this endlessly. Consider the last time you left a tip, recycled your plastic bottles, or gave some change to a street musician. People find themselves in prisoner's dilemmas constantly and yet almost never choose the dominant strategy. We need an explanation for why we cooperate when it is not in our interest, and neither maximizing utility nor maximizing fitness seems to be able to provide one.

Fortunately for humans and other animals, nature rarely imposes single prisoner's dilemmas on interacting organisms. Much more frequently it imposes repeated, or iterated, prisoner's dilemmas, in which each of many individuals must play the game many times, either with the same or different players. For example, every purchase over the counter at a shop one frequents regularly is a single turn in an iterated PD.

Under fairly common circumstances, there is a far better strategy in the iterated prisoner's dilemma than defecting. As Robert Axelrod showed in a series of computer models of repeated prisoner's dilemma games, almost always the best strategy is tit for tat (TFT): Cooperate on round one, and then do what your opponent did in the previous game. If in game 1, or in game *n*, the opponent defected, tried to free ride on your cooperation, then on turn 2, or on turn *n* +1, you should decline to cooperate, you should defect. If on turn *n* + 1, the opponent switches to cooperate, on turn *n* + 2, you should go back to cooperation. The conditions under which Axelrod argued that TFT is the best strategy in the iterated prisoner's dilemma include: (a) there is a nonzero probability of playing the game with this opponent again, that is, if this round is known to be the last game, there is no point cooperating in order to encourage further cooperation; and (b) the value of the payoffs to cooperation in the next games is high enough to make it worthwhile taking a risk cooperating in this game to send a signal that you may cooperate again in the next game, if the other player cooperates in this one. In both computer models and PD tournaments among real players, TFT almost always comes out on top.

For example, suppose we set up an experiment in which 1,000 undergrads play a prisoner's dilemma for money one hundred times with randomly chosen opponents. The payoff for mutual cooperation is, say, five euros or dollars or some other unit of currency large enough to buy, say, a beer; the payoff for defecting when the other player cooperates is ten; the payoff for cooperation when the other player defects is only one unit; and the payoff for mutual defection is three.

Suppose that after every ten turns we eliminate the players whose strategies tied for earning the smallest winnings and increase by the same number the number of players whose strategies secured the highest winnings. This

		YOU	
		Cooperate	Defect
ME	Cooperate	5 II 5	10 I 1
	Defect	1 IV 10	3 III 3

The prisoner's dilemma for a computer tournament.

simulates an environment that filters out the less fit and selects for the fitter PD strategies. We can expect the players to employ a variety of strategies, such as always cooperate, always defect, cooperate till first defection and defect thereafter, flip a coin, cooperate if heads, et cetera. When experiments of these kinds are run, TFT almost always emerges as the winning strategy. That is, in experimental circumstances with the sort of payoffs experimenters can afford to provide, reasonably well-educated people in their late teens (university students) raised in different cultures all over the world, generally find themselves cooperating in the iterated prisoner's dilemma. Likewise, when we program computers to simulate such a tournament over a wide range of payoffs, distributions of alternative strategies, number of turns in the game, and a diversity of weightings of future payoffs to present payoffs, the results come out the same. TFT, the conditional cooperative strategy, wins.

Axelrod explains this result by pointing to three features of this strategy: (a) it is nice—that is, it begins by cooperating; (b) it is retaliatory—it cannot be treated badly more than once without punishing the defector; and (c) it is clear—opponents don't have to play against it many times to figure it out and fall in with its strategy to their mutual advantage. Of course, game theorists have recognized that TFT is not always the best strategy. For example, suppose every player makes mistakes a certain proportion of the time by, for example, pressing the defect button when they meant to press the cooperate button simply by accident once every ten turns. In an environment that contains mistake-prone players, tit for two tats might do better, as it is slightly more forgiving and so will not provoke as many mutual defections caused by sheer accident. Or again, suppose that the set of players contains some significant number of altruists, who always cooperate no matter how their opponents have played against them in the past. Under these conditions, a tat-for-tit strategy of defecting first and then switching to cooperation if the opponent switches to defection may do better than tit for tat. Nevertheless, the approach originated by Axelrod does vindicate conditional cooperation as likely to be the fittest strategy in a broad range of iterated PD situations. But if the simplest way for human behavior to have been shaped, reinforced, and selected for such cooperative dispositions in iterated prisoner's dilemmas is by giving us dispositions that lead to cooperation, it will be no surprise that we pay for our purchases, are not cheated by salespeople, leave tips even when we don't plan to return to the restaurant, drop change into a street busker's hat, or put out our bottles for recycling when it is easier to shove them into a trash bag.

Iterated PD is not the only game in which being nice to others has a higher payoff to the individual player than does looking out for number one. Consider three other games, or strategic interaction problems, that game theorists and experimental social scientists have explored. In cut the cake, two anonymous players are each asked to select some portion of a significant amount of money, say ten dollars or ten euros or some other significant currency unit, on the condition that if the other player's selection and theirs add up to more than ten units, neither gets anything, and if it is equal to or less than the total, each receives what they selected. In this game, almost everyone pretty spontaneously asks for half the amount. Consider a second game, called ultimatum. One player, the proposer, specifies how much of the ten dollars the other player will receive and how much the proposer will keep. If the second player, the disposer, agrees, each party gets what the proposer decided. If the disposer declines, neither party gets anything. In this game it would obviously be irrational ever to decline even a quite unfair split, since even a small portion of the total is better than nothing. And yet, across a broad range of cultures (including non-Western nonuniversity students) in which even one-tenth of the total to be divided in the experiment is a significant amount, parties to the ultimatum game almost always propose a fair split and reject anything much less.

What is interesting about these two experimental results is that the acting on preferences for fair and equal division that each of them reveals has been shown to be the winning strategy—having the highest payoffs in iterated cut the cake and ultimatum computer models that simulate natural selection for optimal strategies by self-interested agents or fitness maximizers. Of course, such results are significant for explaining human commitments to fairness or equality only using a number of important assumptions: the payoffs in the model games must reflect real-life alternatives, the interactions have to arise frequently enough in real life so that players' choices have effects on their future opportunities, and most of all there needs to be some device or disposition in human beings that makes them adopt strategies that are costly in the short run and beneficial in the long run. This is a point to which we shall return.

A third game could be of particular importance for understanding the emergence of teamwork and other multiple-agent, cooperative social practices. It is called stag hunt, after an idea by Rousseau. To successfully hunt a deer requires a group to surround it, but each member may be tempted to leave the circle of stag hunters if the prospect of trapping a rabbit arises. So why would a rational agent even begin to hunt the stag, if the prospect of at least one other hunter's defecting to trap a rabbit will make the whole stag hunt fail? Here is the game:

		YOU	
		Hunt stag	**Trap hare**
ME	**Hunt stag**	4 II 4	3 I 0
	Trap hare	0 IV 3	2 III 2

The stag hunt game.

In this version, if we both go for hare trapping, the result is a smaller payoff than if only one does, reflecting an assumption that there is some costly interference between hare trappers. Here, as in the PD, we both prefer box II to box III but, unlike a PD, each of us prefers it above all other outcomes. Another important difference from the PD is that my best strategy is contingent on what you do. (Recall, in the PD, my best strategy is to defect, no matter what you do.) In the iterated stag hunt, there are several potential strategies, including a version of tit for tat: start out hunting stag and continue to do so with those who hunt stag with you, but trap hare when your potential fellow stag hunters have switched to hare trapping the last time a stag hunt was undertaken. In general, the conditionally cooperative strategies of stag hunting do far better than invariable hare trapping or strategies that hare-trap from time to time or when it appears advantageous. Biological anthropologists will notice that if the stag hunt models the strategic problem facing prehistoric human hunter-gatherers, our Homo erectus ancestors, then it will be no surprise that social cooperation emerged among them long before the emergence of other forms of life, such as agriculture and the social changes it produced. Sociologists will see that the persistence of cooperation need not be incompatible with individual fitness maximization. Economists will ask themselves whether individuals engage in opportunistic free riding in such situations at levels below those at which detection and punishment become worthwhile to cooperators. Psychologists will want to know what human characteristics nature has capitalized upon to set up these patterns of cooperation.

What the evolutionary game theory models show about cooperation, fairness, and equality in behavior is that *they could have arisen* by natural selection operating on strategies employed by individual human agents or, for that matter, by other organisms. But these models by no means show that cooperation, fair dealing, and a preference for equal divisions is, so to speak, *in the genes*! The Darwinian dynamics of removing the less-fit strategies and multiplying the more fit, which evolutionary game theorists invoke, can operate over rounds of play in a tournament just as effectively as over generations of reproduction. It can select for winning strategies and increase their proportions in a population of strategies on the basis of learning and imitation just as well as differential reproduction. In fact, if we add very simple learning rules to these models, they produce cooperative, fair, and equality-favoring outcomes even more reliably and quickly than the pure elimination of the less-fit and reinforcement of more-fit strategies at the end of each round. For example, in the ultimatum game, in cut the cake, or in stag hunt, suppose each player is surrounded by eight neighbors, as on a grid, and we can add to the model the rule "After each round, switch to the strategy employed by one's most successful neighbor." Under these circumstances the strategies that reflect short-term altruism almost always do best in terms of long-term self-interest.

Philosophers, game theorists, and other social scientists have recognized that, for agents who play cooperative strategies in iterated strategic interaction, there may even be a version of group selection that the methodological individualist can accept. In doing so, it would provide another crucial component that the functionalist approach to social theory requires. Consider a group of PD cooperators, or fair-minded cut-the-cake and ultimatum players from which a free rider emerges, either by immigration, mutation, or a calculated switching to another strategy. Provided the individual game payoffs for cooperation are close enough in value to the payoffs for defection, and that the number of cooperative players is high enough, over time the number of free riders will not grow beyond a certain manageable proportion of total players. Once the number of free riders passes this threshold, they will meet one another often enough in iterated games to secure lower total payoffs than cooperators would have. Moreover, only a small amount of learning will be required by players in order for cooperators and those with a preference for fair strategies to seek one another out and interact preferentially, thus increasing the disparity between long-term payoffs to cooperators versus free riders. In groups where the costs to a third party for punishing free riders in two-person games are low enough, and the costs of being detected and punished to the free rider are high enough, there will be even stronger barriers against the subversion-from-within problem we noticed in Chapter 10. Freed from the threat of inevitable subversion from within, the advocate of group selection can begin to secure evidence that conditions for its persistence frequently obtain. And the methodological individualist can accept the mechanism that brings about what the group selectionist calls groups, and what the individualist may prefer to call correlated individual strategies.

Besides their freedom from any need for Durkheim's collective conscience, and their reliance for survival on the aggregation of individual choices, there is one other thing of importance to notice about these groups. Their natural selection and survival hinges on rather sophisticated cognitive feats—recognition, memory, and calculation of costs and benefits—which are difficult to hardwire genetically, and certainly cannot have obtained among infrahuman animals. If this sort of group selection is to be allowed by the methodological individualist, it will be only for groups of organisms of high intelligence. And this sort of group selection, if that's what it is, will be vindicated only if the social scientists who make use of it can show that the strategic interactions of individuals satisfy the payoffs of these games, or ones with the same properties.

FAREWELL TO THE STANDARD SOCIAL SCIENCE MODEL?

Nevertheless, there is a persistent line of theorizing in social and behavioral science that does attribute the emergence of these dispositions to cooperation, fairness, and equality, and to genetic traits that have arisen, been transmitted, and selected for in the same way that almost all other adaptations among biological creatures have been: through our genes. These hereditarians reject what they call the standard social science model according to which the mind is, as the British empiricist philosophers supposed, a blank slate so that most behavior it gives rise to is learned and little is innate, hardwired, or genetically preprogrammed.

The standard social science model, or SSSM, is a label sometimes imposed on the rest of the social and behavioral sciences by exponents of a strongly evolutionary approach to human affairs, an approach originally labeled *sociobiology* by its earliest influential proponent, E. O. Wilson, and later labeled *evolutionary psychology* by social scientists eager to avoid some of the controversy that surrounded Wilson's claims—and also eager to emphasize their commitments to a methodological individualism that made (social) psychology central to the explanation of how and why social institutions have their characteristic functions.

It is fair to use the label *standard social science model* for the majoritarian tradition in social science, especially self-consciously "empirical" scientific studies of human affairs that emphasize the control of theory and explanation by data. Empirical inquiry is, of course, important regardless of whether one advocates a "nature" or a "nurture" theory of human affairs. But the method that involves varying environmental variables either in real experiments or in quasi-experiments that nature may arrange for the observant scientist, seems to vindicate an environmentalist assumption according to which people learn from their environments. For if significant differences in behavior are correlated with differences in environmental conditions, it is widely presumed among social scientists that it is environmental conditions that cause the behaviors, as a result of some process of learning. Hereditarian opponents of the SSSM reject this argument as a simple-minded mistake that fails to reflect two facts. First that the environment can bring about nothing unless it works with what is hardwired in an organism's genetic inheritance, and second, that this inheritance constrains the organism's responses to a very narrow range of outcomes. If this is right, to understand human behavior we have to reject the SSSM and admit that nature has as much as or more to tell us about human affairs than nurture.

A metaphorical contrast expresses the views of these sociobiologist and evolutionary psychologists effectively. In contrast to SSSM, which views the mind and brain as largely programmed by the environment, including, of course, the social and cultural environment, these "nativist" scientists view the mind more like a Swiss army knife. That is, they treat it as a package of a large number of different special-purpose instruments, each with its own independent domain of operation, each the developmental result of a distinct package of genes that were selected for by the evolutionary environment from which Homo sapiens and its immediate ancestors emerged. The brain is composed of functionally specialized *modules* in this view, each of which has been separately selected for over a long evolutionary process.

The notion of a mental module was introduced by Jerry Fodor, a philosopher, in the early 1980s and has become fashionable among those arguing that many of our behavioral traits are genetically hardwired. As Fodor understood them, mental modules engage in processes that are broadly "computational"—they are biological computers, designed to (selected for) efficiently and quickly solve significant problems in very specific domains, by processing only limited amounts of the large quantity of information that may be available to the agent; that is, they are "epistemically encapsulated." Modules are required to learn what the environment has to teach us quickly enough for any of us to survive infancy. Accordingly, they will have to be hardwired in the brain, and therefore cannot be much influenced by environmental information—their epistemic powers are "bounded." One of Fodor's favorite and relatively uncontroversial examples of a module with these features is the part of the brain responsible for visual perception. Given the two-dimensional data available on the retina, this module solves the domain-specific problem of constructing a quite different three-dimensional representation of distances, sizes, and shapes by processing the retinal image quickly and unconsciously, employing an implicit theory of how things look related to how they are. This module operates with a theory that is encapsulated and bounded (whence the visual illusions we can be subject to), but that provides a highly adaptive output almost all the time. It is uncontroversial that the visual system is the separate and distinct result of evolutionary selection for the solution of a pressing design problem. The question in dispute between hereditarians and hereditarian or "nativist" evolutionary psychologists is how much more of the human mind is a matter of innate modules.

The grounds for nativism include at least one very surprising experimental result and one type of general argument. The experimental result is highly relevant to the game-theoretical argument for the emergence of cooperation as an evolutionary stable strategy among fitness maximizers. It is widely recognized that almost all iterated strategic encounters are open to exploitation by a certain amount of free riding—that is, defecting, demanding more than a fair share, hare trapping instead of stag hunting, et cetera. In many reasonable models of these interactions, the cooperative strategy can persist in a stable equilibrium with a modest amount of cheating, or other short-term selfishness. But, under most circumstances, in order to prevent free riding from swamping cooperation over the long run, either the group of predominant cooperators must send out colonies of predominant cooperators with a frequency that depends on the level of free riding in them, or the free riding has to be policed and punished and the free riders shunned in the opportunities for cooperative interaction. What this requires is a *free-rider detection device*, and there is some evidence that we have such a device and that it is hardwired, that is, genetically encoded.

The evidence is in an experiment called the Wason selection test. Subjects are asked to solve two problems that are formally, that is, mathematically identical. Problem 1: You are a bartender and you must enforce the rule that all beer drinkers are above eighteen years of age. There are four persons in the bar. *A* asks for a beer, *B* asks for a lemonade, *C* says she is seventeen, and *D* is obviously elderly. Whose identification card should you ask for to ensure there is no underage drinking? The answer of course is *A* and *C*, and almost everyone gets this question right. Problem 2: There are four cards in front of you, marked *A, B, 5,* and *6.* Which cards must you turn over to determine whether every card with a vowel on one side has an even number on the other? The answer is *A* and *5;* turning over the A is pretty obvious; 6 can have a vowel or a consonant on its reverse without falsifying the claim that every card with a vowel on one side has an even number on the other. If the 5-card has a vowel on the reverse it will violate the rule. Fewer than 25 percent of subjects get this problem right. Logically speaking, the two problems are exactly the same! And yet, even persons who have studied logic do no better than most others on this problem. What is more, this result is cross-culturally robust. Change the problems in ways that make them familiar across a variety of cultures and groups within them—east-west, developed-undeveloped, urban-rural, educated-uneducated, male-female—and you get the same result. For example, make the rule that if you go to Mecca you must be a Muslim, or if you wear a sword you must be a samurai, et cetera, and people will invariably be able to identify who must be checked to enforce such social rules. But right across the same cultures, people cannot similarly solve the logically identical problem in which abstract symbols are substituted for socially significant status markers.

Influential evolutionary psychologists argue that the universality of this finding suggests that people have an unlearned, hardwired, domain-specific, cheater-detection capacity, one selected for enabling them to monitor social interactions for cooperative-rule violations. After all, the only difference between the two problems is the application of reasoning to a problem of cheater detection in a social context, which is absent in the other problem. If, cross-culturally, people perform differently on the two problems, then the cause of the performance difference is probably not environmental, and the cognitive equipment that solves the cheater-detection problem must be different from and independent of general reasoning capacities or whatever we use to solve purely logical problems. Ergo, there may be an innate and hardwired, genetically encoded mental module whose function is to identify those with whom cooperation is profitable and those with whom it is not. And this is just what would make the evolutionary emergence of a disposition to cooperation, fairness, and equality possible.

THE POVERTY OF STIMULUS AND OTHER ARGUMENTS FOR INNATENESS

The theoretical argument for the claim that our behavior is the result of the operation of cognitive models that are hardwired in our brains by natural selection generalizes from one advanced by the linguist Noam Chomsky for the innateness of a language-learning module in the human mind. This argument begins by pointing out the "poverty of the stimulus" on the basis of which young children very quickly learn their first language, and the richness of the linguistic competence they acquire in so short a time. Only a year after birth, most children, regardless of their intelligence or the language to which they are exposed, who have been exposed to a minimum (and perhaps even a highly defective) amount of their caregiver's language, begin to speak it. Soon thereafter they can encode and decode an indefinitely large number of completely novel and utterly different expressions in a bewildering variety of grammatical structures, some of which they may have had hardly any exposure to. It was Chomsky's conclusion that this feat was possible only if children came into the world equipped with a hardwired language-learning device or module, an innately preprogrammed set of rules about language that enabled the child very early, and quite unconsciously, of course, to recognize that some of the noise one hears is a language, and to frame a series of hypotheses about the grammar of that language, which one then tests by one's responses to the linguistic and nonlinguistic stimuli of other speakers. The innateness of a language-learning device is now widely accepted in linguistics and psychology generally.

Chomsky's "poverty of stimulus" argument was so powerful that it spawned arguments for the innateness of a number of other universal human capacities. These arguments claimed there are parallel "poverty of stimulus" and "richness of competence" that would underwrite other attributions of innateness to other human capacities, and thus ground further rejection of the so-called standard social science model. Evolutionary psychologists have sought to explain the early and quick emergence of certain phobias about snakes, mushrooms, and other potential threats to health as reflecting the operation of a "folk biology" module. In light of infant-gaze experiments, they hypothesize a hardwired "theory of other minds" to explain even infants' abilities to attribute motives in human actions. A slightly different argument suggests that there is an innate "folk physics" reflected in the human and, before it, the infrahuman mind. Owing to our long evolution in a world of physical regularities that are very stable and very important to know about, or at least learn about quickly and very early for survival, it will be more adaptive to have these generalizations hardwired than for them to be learned by experience in every generation.

In addition to arguments for the innateness of certain cognitive potentials and abilities that shape behavior, there is also an argument for the innateness of certain emotions and other affective psychological phenomena. The universality of certain emotions and the commonality of their expression by humans and other animals was already noticed by Darwin and reported in one of his last books, *The Expression of the Emotions in Man and the Animals* (1872). In fact, that capacities to feel certain emotions under common conditions are hardwired and genetically encoded is something that evolutionary game theorists need to make their models relevant for the emergence of human cooperation. It is evident that human beings infrequently act cooperatively just from conscious deliberation and calculation, showing the long-term advantage to their own benefit; and people almost reason their way to cooperative strategies by considerations of the consequences of cooperation to their long-term reproductive fitness! People rarely free ride or cheat even when they realize they can get away with it. Accordingly, the fact that cooperative strategies like TFT are fitness maximizing or utility maximizing is by itself insufficient to explain the actual emergence of the fact of or the disposition to cooperation. Therefore, something other than the desire to maximize long-term benefit or evolutionary fitness must be the cause of cooperative behavior.

Now, what evolutionary game theory at most shows (provided its assumptions are reasonable) is that if there is anything in the human (and infrahuman) psychology that causes, or even just encourages, cooperative behavior, there will be selection for it provided that the cooperative behavior is fitness maximizing. What it needs can be provided by a theory of the innateness of emotions. The reasoning here is quite similar to the explanation for the near universality of orgasm among humans. Nature will select for anything that increases reproductive rates; therefore, it will select for anything that increases the frequency of sexual relations. Accordingly, it will select for those animals that find sex pleasurable and therefore engage in it with high frequency. Ergo, any physiology that makes orgasm a by-product of sex will be under strong favorable selection and will become nearly universal quite quickly on an evolutionary timescale. The same sort of argument suggests that an invariable linkage between free riding and aversive emotions, such as feelings of guilt after free riding or other cheating, or between feelings of sympathy and subsequent acts of sharing, or between the emotions of anger and disdain and the disposition to punish free riding will be strongly selected for, owing to their likelihood of encouraging cooperation. Notice that for such emotions to work effectively to encourage cooperation and discourage selfishness, they will have to be difficult to fake or to suppress even when the agent calculates it is advantageous to simulate or suppress them. But, the argument goes, only biologically hardwired emotions that are out of conscious control could satisfy this requirement.

WHY IS THE NATIVIST/SSSM DEBATE SO HEATED?

Encouraged by these and similar arguments, "innatist" or "nativist" opposition to the so-called standard social science model grew steadily over the last few decades of the twentieth century. The debate between these sociobiologists, evolutionary psychologists, and behavioral biologists on the one hand and behaviorists, learning theorists, and environmental determinists on the other has become extremely heated, and has spilled over from the purely scientific to the broader academic arena, and indeed to a wider public. The reason is pretty clear: nativism about socially significant human traits may have profound consequences for public policy, for the persistence and enforcement of social mores and norms, and for people's attitudes and prejudices about others. Nativism explains the distribution of traits as hereditarily fixed—genetically encoded and adapted by a long process of selection to a local environment. Thus it is easy to infer from such explanations that attempting to eliminate such traits will be harmful or impossible. Accordingly, society needs to resign itself to their persistence, whether we like them or not.

Among the earliest encouragements to a nativist approach to social institutions was the research program of explaining the almost universal incest taboo. Once Mendelian genetics came to be combined with Darwin's theory in the early twentieth century, it became apparent that, as a biological practice, incest would be naturally selected against. This is owing to the increased likelihood that the offspring of genetically related individuals will suffer from the expression of recessive fitness-reducing hereditary abnormalities; so any tendency to incest would be selected against. But this conclusion leaves unanswered the question of how nature implements the avoidance of inbreeding. How do individuals avoid electing sexual partners when they cannot detect and do not act on degrees of genetic relatedness, indeed when they do not even recognize the connection between sex and procreation, as with all infrahuman and some human groups? What is the proximate mechanism for the evolutionarily adaptive pattern of incest avoidance we find almost universally among humans and most infrahuman vertebrates?

The most well-confirmed theory of the proximate causes of incest avoidance in the human case is traced to Evart Westermack, an anthropologist at work in the early twentieth century. Westermack's theory suggested that nature solves the problem of incest avoidance by a simple "quick and dirty" solution to the problem of detecting close genetic relatedness: humans have a hardwired disposition to acquire an aversion to sex with any person they were reared with during early childhood. By and large, co-reared children have, over evolutionary timescales, almost always been genetically related (community care of unrelated children is a relatively new and uncommon institution). Thus, Westermack theorized, simply avoiding sex with any co-reared potential partner is a satisfactory solution to the genetic relatedness problem that incest avoidance raises. Evidence for Westermack's hypothesis has continued to strengthen over the past several decades. For example, when unrelated children are reared together, the frequency of sexual relations among them after puberty is well below normal. For another example, note that the fitness risks of inbreeding for females of almost all sexual species will in general be greater than for males, since the offspring of any single incestuous union will be a far higher proportion of the female's total number of offspring than of the males. So the Westermack hypothesis should lead us to expect that females require less co-residence with a potential sexual partner to inhibit sexual relations. Indeed, evidence bears this out.

The difference between potential offspring numbers between males and females is the consequence of a genetically encoded trait of mammals: females have a hundred ova or so, and males have literally millions of sperm. This difference can not only be expected to make a difference for the increased strength of incest avoidance by females, it has also been often cited to explain a variety of social differences between males and females. Thus, the crossculturally common double standard that treats male promiscuity as normal and female faithfulness as the norm has been explained as the evolutionary outcome of differences in adaptational strategies due to the sperm/egg difference. Since mammalian males can impregnate a large number of females, their fitness-maximizing strategy is to attempt to use all their resources to do so, and not to mate with just one female and devote resources to their joint offspring. The universal uncertainty of male paternity adds further force to the adaptational value of this strategy. By contrast, females have only a limited supply of eggs, and their optimal strategy is to seek mates with "good" genes and exchange sexual access for the male's long-term commitment to convey resources to her and her offspring. Recent work by ornithologists has also suggested that when there is a payoff to undetected "extra-pair" copulation with another male fitter than the female's partner, this strategy will also emerge among female birds. According to evolutionary theory, male birds are already independently under selection for participation in such behavior. Additionally, and for the same reasons, among most mammals there will be strong selection for offspring caregiving by females, but not for offspring care by males. Now consider the inferences drawn for the human case from these fairly robust theoretical claims about most mammals and some bird species.

It will be tempting to infer that male marital infidelity is in the genes, and is the result of so long a pattern of selection that there is nothing much that can be done about it. What is more, the persistence of rape and sexual assault may be explained as the expression of this disposition so firmly entrenched by eons of evolution that those males who engage in it should be viewed as victims of their genetic makeup, not really responsible for their conduct, and therefore excused from punishment for it. Thus, the inference from selection for sexual strategies in mammalian males to the explanation of criminal misconduct among Homo sapiens males is said by some to encourage complacency about violence against women: "It's inevitable and all we can expect to do is minimize it. We will never be able to eliminate it."

A similar explanation is advanced for the distribution of gender roles characteristic of most societies, in which females are typically the sedentary, stay-at-home, child caregivers, and the males are the out-of-the-house hunters/farmers/craftsmen/traders, et cetera. These differences are the result of natural selection on traits fixed by heredity and optimal for men, women, and their children. Gender differences in the norms governing courtship, sex, work, and home, as well as the distribution of various roles and responsibilities in society between men and women, can easily be assimilated to the adaptationally explained differences between males and females in most mammalian species. Again, such explanations are likely to encourage the belief that gender differences in socially significant institutions, norms, and expectations are hereditarily determined and have long-term adaptational value for the species. Accordingly, it will be argued by those who approve of conventional gender roles that attempting to change them may have harmful immediate consequences for the mental and physical health of men, women, and especially children, and long-term maladaptive consequences for their evolutionary lineages. These allegedly evolutionarily inherited obstacles to the departure from traditional gender roles advocated by many feminists and others makes the nativist analysis highly controversial.

Of course it is not just traits deemed adaptational that nativism encourages us to treat as genetically inherited. Alleged differences between groups and genders are also sometimes explained as maladaptations. More than once in the past two generations of social science, it has been argued that (a) intelligence is measured by IQ tests, (b) racial groups are genetically homogeneous, (c) the mean IQ difference among different racial groups is statistically significant, and therefore, probably, (d) intelligence is genetically determined and members of some racial groups are on average less intelligent as a matter of nature, not nurture. More recently it has been argued that gender differences in mathematics or spatial reasoning or other cognitive skills are also due to genetically encoded hardwiring differences due to selection. Like the two previous arguments, these also encourage complacency about inequalities in outcome among men and women or persons of various racial groups. If the argument about the genetic basis of intelligence is correct, we may console ourselves that society is a color-blind, gender-neutral meritocracy in which inequalities will reflect real differences in abilities and not discriminatory treatment.

Finally, just as there is a tempting Darwinian explanation for reciprocal altruism, and for the psychological predispositions that make it possible, and an argument for the availability of mechanisms that avoid the suboptimal inbreeding associated with incest, there will be a similar set of considerations that make racism and xenophobia explainable as genetically encoded dispositions. These attitudes may have been selected for in the distant past, and regrettably they remain with us now in an environment in which, though maladaptive, they cannot be quickly or easily eliminated. The explanation here, too, is easy to construct: just as maximizing genetic fitness militates against interbreeding that is too close, there will also be selection against outbreeding, or reproductive relationships outside of a close kin group. And selection against suboptimal outbreeding will exploit available proximate mechanisms that reduce the likelihood of the outbreeding. Among the most obvious such mechanisms will be fear or hatred or other negative emotion toward strangers, and marks, signs, and symbols of high kin relatedness, such as appearance, diet, clothing, language, and religious affiliation. Positive markers of group membership and negative emotions toward people who are obviously different in appearance are the result of a feedback relationship between selection for optimal inbreeding. This process has produced a suite of traits most educated persons deplore but to which a Darwinian argument may well resign them. Racism, xenophobia, and religious and ethnic prejudices are determined by factors beyond our control, they will always be with us, and we should accommodate ourselves to this fact.

This pattern of evolutionary explanations for socially significant traits, as individual or group adaptations that fill or once did fill important biological functions, will be unwelcome to many. In particular, those committed to social change, to reform, or revolution to ameliorate the human condition will seek counterarguments to show such explanations are mistaken. Mutatis mutandis, the evolutionary fixity of the status quo, will be good news to conservatives eager to defend current social arrangements as optimal or, if not, then at least as inevitable.

Many leading evolutionary biologists have counted themselves among those dissatisfied with the social status quo and as eager to ameliorate social problems. This has motivated them to seek arguments against the general research program, first of sociobiology and then evolutionary psychology, for the fixity of psychological traits and social norms. These critics of the application of the theory of natural selection to explain functions as adaptations complain that they may well be telling "just-so stories," too easy to construct and even easier to defend against contrary evidence. But the just-so story critique of adaptationalism was not the only arrow in the quiver of the opponents of nativism, as we shall see in the next chapter.

Introduction to the Literature

The classic argument for functional social science is Emile Durkheim, *Rules of the Sociological Method.* John Elster, *Nuts and Bolts for Social Science,* is an excellent introduction that makes clear the need such theories have for a Darwinian mechanism.

The combination of Darwinian biology and game theory in explaining human cooperation begins with Robert Axelrod, *The Evolution of Cooperation.* Some of the most philosophically interesting modeling of the evolution of social norms and institutions is to be found in two books of Brian Skyrms, *The Evolution of the Social Contract,* and *The Stag Hunt and the Evolution of Social Structure.* Steel and Guala's anthology contains a good summary of some of this work by Skyrms and Jason Alexander, "Bargaining with Neighbors."

Sociobiology begins with a vast work of that name by E. O. Wilson, and is subject to withering criticism by Phillip Kitcher, *Vaulting Ambition,* much of it motivated by moral outrage owing to anxiety about the prospects for genetic determinism it is feared that Wilson's books may encourage. A more accessible work defending his views is Wilson's Pulitzer Prize–winning *On Human Nature*. The two names most closely connected with arguments for evolutionary psychology are Leda Cosmedes and J. Tooby, whose latest work is always available online, at www.psych.ucsb.edu/research/cep/primer.html. Their critics are legion, but often intemperate. A reliable place to begin with criticisms is Fiona Cowie, *What's Within? Nativism Reconsidered.*

CHAPTER TWELVE

Theories of Cultural Evolution

Can social scientists solve the problems faced by their holism and functionalism without adopting a nativist view of human affairs? This chapter explores the prospects for adapting Darwinian theory to solve these problems without committing the sciences that employ the theory to any very strong genetic determinism.

THEORIES OF CULTURAL EVOLUTION

The general thesis that some socially significant traits are largely under genetic control, and appear invariant across a range of environmental conditions of development and expression, is often labeled *genetic determinism*. As noted in the last chapter, it is the consequences of such a thesis for public policy and social attitudes that make many persons seek to refute the claim, not just in its individual instances but as a coherent possibility in general. Since the advocacy of genetic determinism often persists even in the absence of evidence for particular cases, it is tempting to accuse its proponents of ideological and political motives. But such accusations will not put an end to the tendency to attribute socially significant traits more to nature than nurture. And besides, there are certainly many "nativists" who share the ameliorative motivations of their opponents, and who nonetheless believe that the evidence favors some version of genetic determinism for some traits.

BLUNTING THE THREAT OF GENETIC DETERMINISM

One of the strongest arguments against genetic determinism challenges the nativists' notion that it even makes sense to talk of a "gene for *X*," where *X* is a trait of interest to the social scientist. If so, there will be no such thing as the gene(s) for IQ or child rearing or xenophobia, et cetera. To see how such an argument would go, consider the case of phenylketonuria, or PKU, an inborn error of metabolism, which really does look like a genetic defect in which the child's inability to make enough enzyme to metabolize phenylalanine and its buildup in the brain produces mental retardation. Is even this clear-cut case of a hereditary defect one that is genetically determined? Probably not. It is well known that the syndrome can be avoided by a simple environmental manipulation: keep the child away from phenylalanine (look at the label on the next can of a diet soft drink that comes into your hands: "Phenylketonurics: contains phenylalanine"). More to the point here, however, the syndrome of PKU can be produced by point mutations in any of a hundred different loci of the genetic material, the symptoms can also result from mutations in genes for the production of other enzymes needed to metabolize phenylalanine, and it can result even in the presence of normal genetic inheritance if, during pregnancy, the mother either consumes a great deal of phenylalanine or is unable to metabolize the amino acid. So, strictly speaking, PKU sometimes turns out to have an independent environmental source, and not to be genetically determined at all. These difficulties will be vastly multiplied when we move from the effects on early development of the immediate enzymatic products of the genes to the socially significant traits of the adult person, like IQ or a tendency to violence or alcoholism, that "genetic determinism" claims is inherited.

There are other serious objections to the genetic determination of each of the instances of socially significant traits alleged to be genetically determined. Few of the empirical studies that have claimed to show substantial covariation between a behavioral trait, such as a disposition to violence or schizophrenia or alcoholism or risk taking, and a particular genetic locus have in fact been replicated. As for the IQ studies, suppose we set aside questions about the reliability of IQ tests as measures of a single trait properly identified as intelligence, or a small bundle of distinct cognitive capacities, or anything with long-term adaptational significance. Even so, the presumptive failure of the studies of racial differences adequately to control for systematic environmental differences—natural, social, familial—between the racial groups from which IQ data are gathered undermines almost all conclusions about genetic determination of intelligence.

This point about controlling for environmental variation reveals the importance of another determinant of phenotypic outcomes that needs always to be borne in mind when Darwinian theory is applied to social phenomena. Phenotypic traits are the joint products of heredity and environment. And variations in the environment can play a very considerable role in determining variations in the phenotypic outcomes of the same genetic inheritance. Some heritable traits are highly *facultative*: the same genotype will result in quite different observed traits under differing environmental conditions. For example, the same species of butterfly will be dark colored if it pupates in winter and light colored if it does so in spring. And of course there will be no phenotype at all under environmental conditions in which the species is not viable. This means that the trait a gene codes for will have to be understood as specified only relative to an environment. The way in which heritable traits vary in their expression as a function of environmental differences is called a *norm of reaction*. For a quantitative trait such as height we may graph norms of reaction quite simply: treat the y-axis as measuring height and the x-axis as measuring environmental variation, such as annual rainfall or amount of fertilizer used, density of planting, degree of insect infestation, et cetera. Then, the closer to the vertical a norm of reaction is, the more sensitive is the phenotype to environmental variation; the closer to horizontal, the less sensitive is the phenotype.

Now, in regard to almost any socially significant, allegedly genetic trait, it is evident that almost no evidence has been secured about its norm of reaction. And it is obvious why this is so. The difficulty, the cost, and the ethical objections to experiments on the human environment that would be required reliably to estimate the norms of reaction for traits like IQ or schizophrenia or child rearing, are obvious. But without such studies, strong claims that these or other traits are genetic are scientifically irresponsible.

The debate about genetic determinism is of course just a modern-dress version of a much older one, which has traditionally been styled as the debate about nature versus nurture. One problem that has long haunted the debate is unclarity and disagreement about the meanings of the key terms *innate* and *acquired* in biology and outside of it. All parties to the dispute about whether any socially significant traits are innate need to accept the fact that phenotypes are the joint products of the genes and the environment, so that any definition of innateness that could actually apply to any trait will have to accommodate the role of the environment. The same must be said for *acquired*. Acquisition of a trait, say, by learning, requires some capacities, presumably hardwired, to acquire it. Partly for this reason, and because no single definition for either term can be pinned down in ordinary language, and finally because the terms don't appear very frequently in biology (though they certainly do appear in psychology and the social sciences), philosophers have considered whether other terms that are to be found in evolutionary biology do the work these terms have been used to do, and whether employing them enables biologists unambiguously to settle questions of nature versus nurture for various traits of interest.

Interestingly enough, in biology there is no contrasting concept to *innate*. There is no innate/non-innate distinction, and the term has little role in the life sciences. Simple qualitative identity or similarity from generation to generation—"breeding true," for example—may work for the immediate traits of replicators like genes, but not for individual organisms like humans. To begin with, there is the norm-of-reaction issue. When the norm departs much from the vertical, there will be no similarity of traits. If heritability is defined in terms of correlation of parent and offspring traits instead of identity, then we will have to treat many traits clearly produced by nurture as innate. Consider the correlation between parents speaking Hausa and their children doing so. Surely this is not a matter of hereditary determination. Population biologists define high heritability as a ratio that reflects the total amount of phenotypic variation that is due to genetic variation. If the fraction Vg/Vp, genetic variation/phenotypic variation, is close to 1, the phenotypic trait is highly heritable. But, as Andre Ariew has noted, for human populations in which the possession of opposable thumbs is 100 percent, there is no variation. Hence, for most human populations the heritability of opposable thumbs takes on an undefined value. As such it could hardly do the work that innate is supposed to do. Surely, opposable thumbs are innate traits among Homo sapiens, if ever there were any.

Recent work by philosophers and biologists has canvassed at least two dozen alternative definitions for the terms *innate* and *acquired* to be found in the scientific literature, quite apart from their uses in nonscientific contexts. It is no surprise, therefore, that debates about the innateness of traits persist, even when substantial relevant evidence has been agreed upon by disputants. Whether an agreed-upon set of definitions is possible remains a philosophically open question.

MOTHER NATURE OR MOTHER NURTURE?

Is the only alternative to genetic determinism just to repudiate the relevance of Darwinian natural selection to human affairs? There had better be another alternative. For as noted several times now in the last three chapters, we need Darwinism if we are to give any functional explanation of human behavior at all! Too much human behavior, and too many human institutions show every mark of having functions, some manifest, others latent, for this fact about them to be disregarded and left unexplained. Those social sciences that appeal to functions, especially to latent ones, require some mechanism or other short of final causation to bring about and to maintain the functions these disciplines identify. But there is only one such mechanism.

What social and behavioral science requires can only be provided by Darwinian natural selection. So the question becomes whether there can be such "natural selection" without genes, without the nativists' commitment to Mother Nature as the explanation of behavior?

There are both social scientists and philosophers who accept Darwin's theory as the only available account of functional traits but who reject nativism. They have worked hard over the past three decades to construct Darwinian theories of cultural, social, and psychological evolution that are free from a commitment to exclusive or even substantial hardwiring, innate modules, and the genetic determination of traits that nativists argue for. All these nonnativist Darwinian theories of cultural evolution require of the human mind/brain is that it be especially good at learning by imitation. On this view the only innate, hardwired modules are the small number of sensory input devices such as sight, smell, touch, taste, et cetera. These alternative Darwinian theories don't treat the mind as a sort of Swiss army knife set of domain-specific, epistemically encapsulated and bound modules. Instead of selection for genetically encoded modules, they hold that it has been cultural selection for traits that are not genetically encoded, selection for packages of adaptive information preserved and transmitted outside the brain.

Biological anthropologists and philosophers of biology have developed scenarios for human evolution in which environmental and learned psychological factors work together persistently to produce many of the adaptations the evolutionary psychologist is tempted to explain genetically. The most important of the psychological factors they identify as crucial to human cultural evolution is the very great developmental plasticity of our brains—and its consequent remarkable powers of imitation learning. Second, they emphasize the strong early influence on our evolution of cooperation as modeled by the evolutionary game theory sketched in the last chapter, but learned and transmitted culturally and not through the development of genetic cheater-detection modules. Perhaps most important, these cultural Darwinians argue that imitation learning and cooperation allow for the construction of relatively long-lasting niches into which subsequent generations are born. These long-lasting niches accelerate the transmission and the accumulation of adaptive strategies for coping with the environment, including technological and cultural strategies. They do this because over time the long-lasting cultural niche in which childhood learning takes place increasingly fosters acquisition of the most efficient learning strategies—that is, the ways in which to filter variant practices for the ones that confer greatest fitness. These cultural Darwinians argue that human neural plasticity is great enough to make for the quick and accurate learned solution to several of the alleged poverty-of-stimulus problems nativists advance.

As we saw in the last chapter's discussion of evolutionary game theory, persistent human cooperation requires individuals to be able to correlate their strategies—tit-for-tat players need to be able to find other tit-for-tat players in order to outperform free riders. This is where intergenerational niche construction—carried on and cumulated over generations—comes into play. For it makes possible the development of learned techniques of free-rider detection and learned enforcement of punishment strategies, along with other strategies that are essential for survival or fitness enhancement. Suitably structured, operating over long childhoods, cultural niches can solve the poverty-of-stimulus problem for all but the linguistic competency that Chomsky first introduced it to deal with. Once human cooperation gets a start, its combination with highly accurate imitation learning by children will also make possible the emergence of cumulative technological change and the specialization of labor, cognitive and otherwise, along with group-strengthening xenophobic traits that are transmitted culturally and will not require a genetic basis.

Clearly, culture and cultural transmission needs cooperation, and therefore requires the emergence of a range of different patterns of reciprocation. For example, unless individual teachers receive a fitness payoff for passing on their learning, lore, and skills, there is no fitness benefit to teaching nonkin some technological innovation (such as toolmaking). So, no cultural evolution without group selection. And the entire process has to get its start from, and continue to be powered by, increasing plasticity in brain development from our primate ancestors through our common ancestors with Homo erectus. For this is what makes for the biggest difference between us and our primate cousins: at some point in the evolutionary past, we became much more adept at learning by imitation than other primates, and this single cognitive development, not a suite of separate modules, may have sufficed for the cumulating reliable transmission of learned, acquired skills, knowledge, norms, et cetera.

Many of the cognitive tasks that evolutionary psychologists believe cry out for a hardwired module can be accomplished by a genetically unmodularized mind, provided that it is smart enough to quickly assimilate adaptive information already produced in previous generations. Exactly what "smart enough" comes to is a matter for debate among several different theories of cultural evolution. This debate makes contact with experiments and computer simulations in evolutionary game theory. Consider three different learning rules that humans are smart enough to adopt, and that can easily be modeled in computer tournaments to see how well they lead to cooperation.

One rule is to copy the behavior used by a majority or the largest number of others; a second rule is to copy the behavior of the highest status individuals; a third rule is to copy the most successful behaviors among those the individual encounters. Each will lead to rapid adoption of the most beneficial strategy under some circumstances. Individuals smart enough to switch among them optimally will do even better, and all without the need for hardwired mental modules.

Deciding who is right between the nativist evolutionary psychologists and theorists of Darwinian cultural evolution will require creative experimental studies. Meanwhile, as noted above, the crucial thing here is that the dispute is a disagreement within a Darwinian research program. The disagreement here is carried on between those who hold that adaptational traits are transmitted genetically and those who hold they are transmitted culturally. Both sides agree on the crucial role of blind variation and environmental filtration in the creation and local adaptation of cognitive capacities. Agreement on a common Darwinian core in any theory of human psychological traits is mandated not only because humans are, after all, biological systems, but even more because when it comes to psychological traits of ours that are obviously functionally adaptive, there really is no alternative to a Darwinian account of their origin and persistence.

CHALLENGING DARWINIAN CULTURAL EVOLUTION

Accordingly, a Darwinian social scientist is committed not just to a figurative or metaphorical application of Darwinian natural selection to the explanation of cultural processes. The applicability of the theory is forcefully denied by many social and behavioral scientists, especially interpretative social scientists, who see human affairs, human action, and human values as an exception to the natural realm, as not governed by the slow process of gradually accumulated adaptation through randomness in the variation of inert macromolecules and passive filtering by the environment. This repudiation of the relevance of Darwin's theory, however, will need hard argument, not just wishful thinking. They will have to show that the necessary conditions for the operation of natural selection are just not present in culture. One way to do this is to demonstrate that nothing like the conditions that make for evolution by natural selection in the wild are to be found in human culture. Therefore, there is no scope for a literal application of the theory.

Recall from the beginning of this chapter what these conditions are: There must be hereditary variations with fitness differences. If these conditions obtain, the theory implies, there will be selection for random variations that are adaptive and selection against random variations that are not. So one obvious objection to a Darwinian account of culture begins by identifying persistent and widespread cultural traits that no one would consider to have functions for, or to be adaptive for individuals or groups. These traits, like tobacco smoking or foot binding, have nevertheless persisted in human society for all of its recorded history.

As we saw in the last chapter, however, this argument that evolutionary theory can't accommodate the long-term persistence of reproductive, fitness-reducing, cultural maladaptations fails to distinguish two distinct processes of selection. Once the distinction is made clear, its force is much reduced. A Darwinian theory of the evolution of culture as the selection of adaptations has to deal with two kinds of adaptational issues. First, it has to account for the evolution and persistence of culturally acquired traits that confer adaptational advantages on individuals. Second, it has to explain how traits that may make no adaptational contribution to individual or group fitness nevertheless persist over time because features of the traits themselves enhance their own reproductive fitness—that is, the frequency with which they are copied by individuals and groups to whose adaptation they make no positive contribution.

A couple of examples will illustrate these two quite different evolutionary scenarios. Consider the disposition to wash hands in hot water before eating versus the disposition to ingest heroin. The former trait confers an adaptational advantage under most circumstances to persons who acquire it, in spite of the costs it imposes. Making a fire, collecting the water in a container that can be heated, controlling the heating to prevent the water from getting too hot, drying one's hands on clean material, and so on are all certainly barriers to the spread of hand washing. Nevertheless, in the long run, the frequency of hand washers will increase. The trait is transmitted quickly and accurately to their offspring, and in the long run the trait spreads in the population, because the lineage that engages in hand washing increases its proportion of the total population, and presumably the total population size increases as well. Washing hands in hot water is fitness enhancing.

By contrast, consuming heroin is plainly fitness reducing, as it has many unfavorable effects on reproductive fitness. On the other hand, the trait itself can spread like wildfire in a population. Once introduced, the practice will be copied, replicated, and parasitize more and more people in spite of the fact that it lowers the reproductive fitness of those who adopt it. Owing to its addictive properties for creatures like us, the practice can in effect parasitize Homo sapiens. Provided that its virulence is sufficiently moderated, its incidence among potential hosts will increase over time until it becomes fixed in the population, crowding out and making extinct other practices, for example, hand washing, child care, and sex, for that matter.

So an evolutionary theory of culture needs to account for cultural traits in terms of their effects on human fitness, and on their own immediate fitness. Failure to make this distinction will prevent an evolutionary theory from explaining the distribution and persistence of cultural traits that are harmful to people and groups but have features that enable them to replicate at higher rates than less harmful traits. Failure to draw this distinction will of course also lead opponents of a Darwinian theory of culture to cite as counterexamples to its claims the persistence of maladaptive traits, that is, traits that reduce people's fitness, when in fact such traits will be as commonplace in a culture driven by Darwinian processes as parasites are common in the biological realm.

But drawing this distinction throws into sharp relief another, graver problem that a Darwinian theory of culture faces. If fitness and adaptation are to be understood in terms of their impact on reproduction, as Darwin's theory requires in biology, it is hard to see how the theory ever could be applied to explain cultural evolution. It is not plausible to suppose that the distribution of cultural traits, which increasingly come and go within the span of a single generation, could possibly be controlled by their impact on human reproduction and the genes humans carry. To do the work required to explain cultural evolution, natural selection is going to have to operate within individual human generations—indeed, within periods much, much briefer than the twenty years a generation lasts—and it will need a mechanism of heredity that operates nongenetically within these brief time periods. And it looks like literal Darwinian natural selection within such time frames will require something else: it will require something that, like the gene, both carries faithful copies of the traits to be selected and allows for continual production of small variations in those traits. Without faithful copying, descendant traits will not be similar enough for their environments to have the same effect on them and thus on their distribution in a population; without random variation, they will not explain how cultural traits are transformed over time. Can the defender of the literal application of Darwin's theory to the social sciences satisfy these requirements? Does the defender of Darwin really need to?

Some who support using Darwinian theory to explain cultural change have taken this two-part challenge quite seriously. They have introduced a concept explicitly modeled on the gene to do the work of applying the theory to culture needs: the *meme*, introduced by Richard Dawkins (in *The Selfish Gene*) and defined as "a unit of cultural transmission, or a unit of imitation." This is not, of course, a very helpful definition. But then neither was the initial definition of *gene* as whatever factor or element in the germplasm results in those traits that are distributed in Mendelian ratios. In the discussion of memes to follow, we need steadily to bear in mind that the time lapse between the introduction of the concept of *gene* and its macromolecular vindication was about fifty years, during which time, the term *gene* was subject to ridicule and skepticism, before it came to be taken seriously as an essential explanatory concept in biological evolution.

The rough idea of a meme is something in the brain that causes behaviors, or some feature of behaviors, and that not only recurs in one brain. Owing to the attractiveness to others of its behavioral consequences for the agent, it is contagious and is copied from brain to brain and results in more copies of the behavior or its features that the meme produces in the first person. Examples of memes are easy to come by. Dawkins's illustration was drawn from animal behavior: bird songs are often not hardwired but copied from generation to generation and in these cases exact copying is critical to fitness. The song is, of course, stored in the brain of parent birds and literally copied by the brains of their offspring. But it is easy to identify memes in human culture: an advertising jingle you can't get out of your head and that you sing aloud, thereby bringing the jingle back into your mind and transmitting it to someone else, who also sings the tune; an idea about how to dress, which results in someone's dressing a certain way and which catches on as a fashion; a way of pronouncing words or making gestures or, more enduringly, the Arabic system of numerals, or the base-10 system of arithmetic, or the rules of addition . . . in short, anything that can be recorded and stored in the brain and that increases or decreases its replication in brains owing to its effects on them. Thus, some memes will be ephemeral, such as "23 skidoo," an expression from the 1920s in the United States, and others long-lasting, such as the "To be, or not to be" soliloquy in *Hamlet.* Among vigorous exponents of the utility of this concept in the explanation of cultural change, Daniel Dennett offers, as an example of memes, *words* that we think of, speak, or write. In this he follows Darwin, who wrote in *The Descent of Man* that "the survival or preservation of certain favored words in the struggle for existence is natural selection."

It is evident that memes can be ideas, thoughts, beliefs, desires, mental images, formulae, theories, languages and their parts, thoughts about music, art, crafts, farming methods, moral norms, spelling mistakes, board games and field games, swimming strokes, the design of artifacts, dance steps, and all of the adjectival modifications of these and other abstract contents of the mind that result in particular behaviors. So whatever the material in the brain that realizes memes, they will have to be multiply realized and heterogeneous in neural structure, location, and relation to behavior.

And, just as in molecular biology we do not identify genes by their DNA sequences but by their functional roles, mainly by the proteins and RNAs they produce, so we should expect to identify memes, if there are any, by their effects in behavior, not by any neural fingerprint. This will, of course, make it much harder to identify and trace the spread of memes in a population, or for that matter their mutation and selection, than it has ever been for genes.

This is where the skeptics about memes and about Darwinian cultural evolution will dig in their heels. Skeptics about memes argue that although memes may be heterogeneous in their neural structure, they will all still apparently have to have one thing in common if they are to provide the basis for a Darwinian theory of cultural evolution. They must replicate accurately; memes in different heads must be accurate copies of the memes in other heads that brought them into existence. The copies of any one meme will have to be countable the way genes are, and there will have to be some basis for distinguishing an instance of one meme-type from the instances of a different meme-type. We will have to be able to count the number of instances of a meme copied from another instance as reflected in the behaviors of the individuals whose brains instances of a given meme-type inhabit. For without such criteria of individuation we will be unable to tell whether a meme has replicated or, if so, whether its fitness is increasing or decreasing, and we will be unable to distinguish memes from one another to decide about their competitive and cooperative relations with one another, or dependencies on one another. In short, the argument goes, unless we can be confident about replication and reproduction, memetic evolution by natural selection will be, as we have said, a mere metaphor, not a scientific hypothesis worth taking literally. But there is no evidence that anything in our minds behaves the way memes will have to work to do the job for cultural evolution that genes do for biological evolution.

How serious is this objection to the existence of memes? Compare the situation for genes. We can no more count memes by observing behaviors than we can count genes by observing intergenerational similarities between plants or animals. To actually identify and count genes from generation to generation required the discovery of laws of genetics, by Mendel, and their refinement over a century. Even using Mendel's laws to identify genes from the distribution of phenotypic traits depends on the precision and the confirmation of other additional hypotheses about the generation-by-generation distribution of various traits in large populations under a variety of environmental conditions. Presumably, the generalizations of this sort will be much harder to establish in the case of human culture than in the case of pea plants! There are several alternatives here: first, it may be the case that, under the welter of behaviors that reflect cultural learning to some degree or other, there is a core of memes that are copied accurately enough, recombine, and mutate sufficiently often to provide the needed substrate for the operation of Darwinian selection in cultural evolution, and that we can in fact discover them! This is the most optimistic scenario for the exponent of memetic-cultural natural selection. A second, less optimistic scenario for meme theorists is that the operation of natural selection on memes does in fact determine all or a great deal of cultural evolution, but most of the memes are too difficult for us to identify and the process is too complex a matter for us to discern in the booming, buzzing confusion of cultural change. A third, still less optimistic alternative is the possibility that there are memes but their mutation rate is very high, even higher than the rate at which an AIDS virus mutates. As such, social evolution will be characterized by a little adaptational natural selection and a great deal of mutational drift. This alternative will not, of course, explain the apparent functional adaptation of so many social processes and social institutions. So, even if right, it will not be of much interest among social scientists. Of course, if there are some areas of social life in which there is reliable memetic copying and a reasonably moderate rate of evolution, there will be scope for a Darwinian theory. But it will not be a generally accepted account of adaptation everywhere in social phenomena.

Then there is a fourth and even less controversial, but also less interesting, possibility: the application of natural selection operating on memes to explain culture is a useful and suggestive metaphor at best, but not a general theory of cultural evolution with anything like as much going for it as Darwinism has in biology.

But the discussion of memes has proceeded on the assumption that the application of Darwinian natural selection to culture really does require a replicator that copies itself accurately in the way genes do. And not all exponents of the literal application of Darwinism to culture will grant this claim. There are important exponents of a literal application of Darwinism to explain cultural evolution who do not think the particulate inheritance that memes would provide is required, and so who do not think their program is hostage to the existence of something in culture that closely parallels the gene. Accordingly, to theorists like Peter Richerson, Robert Boyd, and their collaborators, the items subject to Darwinian natural selection in the evolution of culture need not be much like genes at all. Richerson and Boyd doubt that what they call cultural variants are digital, particulate replicators, largely because they do not think there is much of anything that is transmitted as intact, fair copies from mind to mind. They agree with meme skeptics that there is too much ambiguity in behavior for many different people to internalize exactly the same meme-type by observing a single bit of behavior. But, they argue, Darwinian cultural evolution doesn't need high fidelity copying that plays the role of mental genotype to the behavioral phenotype. Instead they hold that there is a continuum from more to less exact copying that is harnessed together with a range of more or less restrictive learning rules. Recall the three mentioned above, which are easy for humans to employ and equally easy to model in computer simulations that add learning to speed up the selection of cooperative strategies. Intelligence—that is, adapted use of the right learning or imitation rule in the right circumstances—can make up for a good deal of the variation, the ambiguity, and even the mutations conveying memes or anything like them from one person to another.

By employing cognitive heuristics, which independent evidence shows that people employ, large-enough populations of people can share exactly the same trait long enough for it to be selected for or against in the social or natural environment without anything being accurately copied from any one of them to any other of them!

Cultural natural selection theorists like Boyd, Richerson, and their coworkers argue that the existence of replicators and interactors is not necessary but only sufficient for natural selection, that heritable variations in fitness are just one way in which natural selection can proceed. More orthodox Darwinians may call attention to the role in these theories of the psychological machinery that homogenizes disparate cultural variants into classes of behavior similar enough for uniform selection. They may argue that such machinery is both selected for genetically and necessary for the appearance of cultural adaptations through the operation of natural selection.

Perhaps more important, however, than these details is the question—raised by the controversy about memes and copying—of what is necessary as well as sufficient for natural selection as an explanation for cultural adaptation. The questions are crucial not just for explicit exponents of Darwinism in the human sciences but for the use of functional analysis in all the social sciences, something we have seen is almost impossible to dispense with. For as we have seen, Darwin's theory of natural selection provides what is perhaps the only theoretical underpinning available for functions in the social sciences.

SOCIAL SCIENCE AS LIFE SCIENCE

Accepting that humans are biological creatures inevitably drives the empiricist and naturalist to the exploitation of available biological theory in the development of social science, and this attraction strongly motivates interpretative and other nonnaturalistic philosophers of social science to attack the social research inspired by Darwinian theory. The occasionally excessive enthusiasm of the former and the frequent misunderstandings of the latter have made it worthwhile to lay out some of this theorizing, its forms and limits. But there is still another reason to take an interest in the bearing of biology on human behavior, action, and institutions: it may enable us to settle some of the outstanding questions in the philosophy of social science, especially those about causation, explanation, laws, and statistics.

Recall the brief discussion of causation in Chapter 2 ("Progress and Prediction"). There we noted the empiricist claim that causation requires laws. The argument for this claim rests on the view that nothing distinguishes between all causal sequences and all accidental ones except that the former exemplify laws and the latter reflect mere coincidence. One trouble with this claim is that in ordinary life and in many scientific contexts we make true and well-justified causal claims without knowing the laws that stand behind them. Flipping a switch causes the light to go on, but few people can identify all the laws that connect the flipping to the illumination. And no one knows the laws that stand behind such complex causal truths as, "The surprise attack on Pearl Harbor caused the United States to declare war on Japan on December 8, 1941." These facts have led some philosophers and others to suggest that the social sciences need not search for laws in order to justify their causal claims. Rough-and-ready generalizations—"Countries usually declare war when attacked"—will suffice to justify causal judgments. That is, of course, something everyone grants.

However, empiricists and naturalists attach importance to the quest for something better than rough-and-ready generalizations. They traditionally insist that social science needs to secure real laws or at least better approximations to them if it is to improve its predictive success. The trouble is that the causal factors that are individually necessary and jointly sufficient to bring about a particular effect will usually be so numerous that we cannot uncover them. But listing all the factors that are necessary to bring about an effect is just what is required to state a law. That is difficult even in physical science. After all, consider the factors we need to add to, say, striking a match in order to get the "law" out of a ceteris paribus generalization like, "Other things being equal, a struck match will light." We need to add that oxygen is present, the match is dry, the proportions and distributions of phosphorus and sulfur are correct, the tensile strength of the matchstick is greater than the force of the striking, the striking force is great enough to produce a spark, the striking material is dry and produces sufficient friction, and so on. We need to acquire all this information to improve our predictions about match lighting beyond the level of everyday accuracy. Indeed, to improve these predictions we need to eliminate the "and so on" clause, the implicit ceteris paribus clause, from our generalization. In physical science we appear to be able to come close to doing so, once we add clauses about the tensile strength of wood and the friction of striking materials, the minimum composition of phosphorus and sulfur that will ignite, and other relevant facts. Reducing the scope of the “other things equal,” ceteris paribus clause is a large part of the task in learning how to manufacture reliable matches.

Can we expect to do the same thing in the social sciences? Alas, no. A full list of the causes for any general type of social phenomena will probably be complex beyond our ability to grasp. The list will also be heavily disjunctive. That is, it may identify many different alternatives, each of which is sufficient to bring about the same type of effect. Indeed, it will be unlikely that we could ever complete any list of conditions sufficient for any outcome in human behavior, human action, or their aggregations. Our generalizations will always carry ceteris paribus clauses.

Empiricist social scientists have never allowed these difficulties to excuse them from the search for generalizations, laws, or wider explanatory models and theories or the search for useful approximations to them. This is what has made the recourse to statistical regularities of such great importance in social science. A statistical regularity that stands up over a course of observation or, even better, through a series of experiments may function as a surrogate for an unobtainable law. A reliable statistical regularity will underwrite a causal claim and figure in its explanation as well. More important, without at least statistically reliable generalizations, few public policies aimed at attaining desirable social ends could be crafted or implemented.

Statistical hypothesis testing is designed to help us decide whether a statistical regularity reflects the operation of causal forces or factors, as opposed to mere coincidences, mere accidental sequences. Such testing aims to tell the difference between an accidental regularity like “80 percent of US presidents elected in a year ending in 0 die in office” and a causal regularity like “80 percent of smokers in France are unmarried, unemployed, male, university-educated Protestants under the age of forty-five.”

The statistical generalization on smoking is silent on the physical health, wealth, and mental states of unmarried, unemployed, male, university-educated, Protestant French smokers. Does the generalization hold for athletic, overweight, or poor, French, unmarried, unemployed, male, university-educated Protestants? Were 80 percent of the athletic or overweight or poor French smokers also unmarried, unemployed, male, university-educated Protestants under the age of forty-five? This discovery would increase our confidence that the original generalization identifies a causal connection between smoking and those traits. It would suggest that being athletic or overweight or poor were not (significant) causal factors along with age, marital status, education, employment, and religion among French males for smoking. Similarly, if marital status made no difference, we could drop that variable from our generalization. Statistical testing is designed to show whether the statistical regularities we uncover reflect genuine causal relations that could explain them or are just accidental or coincidental relations among social facts.

Once we uncover a relatively well-supported generalization such as our "80 percent of French smokers are unemployed, university-educated males," we would like to do two things: we would like to improve the percentages in the direction of 100 percent by adding further conditions to the generalization (which mining additional statistical data should give us), and we would like to explain the ever-improving generalization by deriving it from some broader theory.

But, try as we might, a statistical regularity such as "80 percent of smokers in France . . ." may not be improvable into a generalization about 90 or 95 or 99 or 100 percent of French smokers. Why not? Can there really be an infinite number of factors that bring about smoking among French people that can be expressed as a complete, 100 percent law as in physics ("All metals are conductors")? Few social scientists will offer this as a reason for the persistence of statistical regularities and ceteris paribus generalizations in social science. But even fewer will give another reason, such as simple indeterminism. And fewer still will argue that we really need higher percentage generalizations or that a reliable statistical generalization about 80 percent of French smokers cannot be explained by more general theoretical considerations.

Once such reliable statistical generalizations are established, social scientists go on to seek explanations for them in mathematical models of the sort rational choice theorists, econometricians, mathematical sociologists, demographers, and others with mathematical tools are familiar with. Usually they seek robust mechanisms that explain the regularities under a wide range of initial or boundary conditions. For a reliable statistical regularity in effect describes what is likely to happen under a relatively wide range of unknown conditions. Theorists argue that these models will explain the statistical generalizations when their assumptions are close enough to the truth to account for the statistical regularity. Others may require that they provide novel predictions about the phenomena they purport to explain or about other processes that social scientists can observe to test the model's predictions.

As we have seen, interpretationalists and other social scientists will cite the difficulty of securing laws as part of their argument that social science has little or nothing to do with generalizations of any kind, or even with causes. When causes are cited, as in the case of history, their being causes is ancillary to their function, which is to help interpret phenomena, to make them intelligible. For this reason neither laws nor statistical generalizations figure in narrative history.

So where does a biological approach to human behavior, action, and institutions come into this debate? Practically everywhere from beginning to end, and everywhere it helps us identify the real limitations on social science and how to adjudicate the debate between empiricists and intepretationalists in a way that shows where each is right in their methodological demands and where each goes too far.

To begin with, some philosophers and biologists have generally concluded that there are really no further laws in biology beyond the three very broad principles of the theory of natural selection: there are hereditary traits; there are variations in hereditary traits; and there are differences in the fitness among these variants to their environments. Everything else in biology, from paleontology to molecular biology to genetics, is the working out of these three laws on Earth over the past 3.5 billion years. And during this period of 3.5 billion years, each "design problem" that one species or lineage of organisms has solved has become a new design problem to be solved by other species who compete with the first species for the limited resources on the planet. The result has been a 3.5 billion–year history of "arms races" in which each adaptation that has been selected for has resulted in new opportunities for the emergence of some new adaptation to exploit it, take advantage of it, turn it into a maladaptation. This cyclical process has in effect made it impossible for any generalization about the traits of a species to remain true throughout the evolutionary history of the planet. Take any generalization of biology you like and there will be exceptions to it, or if there are not yet, the theory of natural selection tells us there will be or else the species will become extinct before the exception arises. Even so fundamental a generalization as the so-called central dogma of molecular biology, that the direction of genetic information is always from DNA to RNA to protein, has been discovered to have exceptions—RNA retroviruses, prions. Not all zebras have stripes, and if lions learn to distinguish stripes from blades of grass, soon enough no zebras will have them or there will be no zebras.

Is there a biological generalization true of all humans, a 100 percent statistical regularity? No, even a statement as basic as "100 percent of humans have twenty-three pairs of chromosomes" is falsified dozens of different ways. Will biologists be able to list all the ways this generalization can be falsified? No. Owing to the arms-race character of interspecies competition, down the road there may be some design problem faced by our species—some new virus or bacterium, for instance—that will provide a new opportunity for selection to vary the number of our chromosomes.

Now, recall the argument advanced in this chapter that wherever biological traits, processes, and behaviors show adaptation to their environment, that is, have functions, the causal mechanism producing them can be only some version of natural selection. Once we exclude future purposes and prior heavenly design, there is no alternative to Darwinian mechanisms for bringing about and maintaining adaptations. And let's apply this idea to behavior, human and otherwise, that is adapted to its environment. As we saw in Chapter 5, it was the behavioral psychologist B. F. Skinner who recognized that natural selection shapes learned behavior just as it shapes genetically encoded behavior. His law of effect is simply the operation of natural selection within the life span of an organism. Of course both the law of effect and Darwinian natural selection require underlying mechanisms—the Darwinian points to genes, the latter-day Skinnerian hopes to identify neural mechanisms—perhaps memes, perhaps something else. And of course if our conscious and nonconscious thinking processes are adaptational (and surely they must be—adapted to solving problems, encoding thoughts in intelligible sounds, protecting ourselves from threats, getting other people to cooperate with us, et cetera), they too must somehow be the products of natural selection, this time operating inside our brains on thoughts that are selected for their fitness to our internal cognitive environment.

As we know from the biological case, however, every adaptation is the solution to a design problem that faces a species, or an individual, or a brain, or for that matter a set of individuals coordinating their actions, or an institution, a strategy, a norm. The result is twofold: first, evolution, and second, the impossibility of those 100 percent regularities that social scientists may seek. As we have seen in the previous section, a behavior like smoking may be selected for, not for the adaptation it confers on smokers but because it is a behavior individuals enjoy in spite of its maladaptive character for them, and so it can persist among certain classes of individuals. Notice, however, that like any biological generalization, this statistical regularity about French smokers, even if it reaches a statistical frequency approaching 100 percent, is subject to change owing to the arms-race character of natural selection. Forces that will change this regularity are easy to enumerate. For example, French smoking may be extinguished by the extinction of French smokers, or by the emergence of a competing behavioral habit, or by an environmental change such as the emergence of antismoking regulations that prevent French people from smoking, or indeed by the internalization among French smokers of memes or similar cultural variants about smoking and health, or the social unacceptability of smoking, et cetera.

If the regularities in social sciences are about traits with functions for humans or ones that parasitize human organisms because, like selfish

DNA, they have been selected for capitalizing on other human traits, it will be no surprise that there are no general laws in the social sciences, just, at most, contingent regularities obtaining for greater or lesser periods of time. Some of them may obtain for decades, centuries, and millennia, such as, "Village commons are likely to be enclosed and privatized by large local landowners," or "Democracies almost never engage in aggressive wars with one another," or, for brief periods, "95 percent of Western teenagers with Internet access download music files." Some may obtain for all human history, such as "Small groups do not need coercion to provide themselves with public goods."

These regularities will themselves be explained in the way that biological regularities are explained: by showing that each of them realizes to a greater or lesser extent a purely mathematical model. In the case of regularities about traits that are genetically encoded, the models that evolutionary biology invokes will be ones about fitness maximization in multigenerational populations, the models of population genetics in particular. In the case of most or perhaps all of the social sciences, the same models will work for regularities that obtain over multigenerational periods, with the understanding, of course, that the traits are not coded for genetically but culturally. In the case of much more temporary generalizations, which obtain with generations or even shorter periods (for example, the one about downloading music), the relevant mathematical models will be those of rational choice theory or, more frequently, evolutionary game theory.

The more we understand the quantitative and the empirical social sciences as biological ones, as disciplines that explain and, to the extent biologically feasible, predict the traits and the behaviors of one particular biological species—Homo sapiens—the clearer and more obvious become the differences between social science and the picture of science derived from physics and chemistry. What is more, the better we can understand the attractions and the indeterminacies of interpretative social science as well. Recall the criticism often lodged against evolutionary explanations that they are just-so stories: too easy to contrive, too difficult to test, and therefore without much empirical content. The search for explanations of why the African rhino has two horns and the Asian rhino has only one leads one to frame hypotheses about alternative packages of environmental design problems and genetic endowments that should have produced each of these traits. These hypotheses are, in fact, "interpretations" of available evidence in terms of varying combinations of environmental and genetic factors, rather like the varying combinations of beliefs about environments and preferences about alternatives that interpretative social scientists advance to make social phenomena intelligible. The evolutionary biologist attempts to make biological traits intelligible in the light of natural selection. And as evidence comes in over time from field ecology, from paleontology, and from gene sequencing, different packages of environment + genes are more or less well supported. But it is probable that the full evidence that will explain this difference is lost in evolutionary history. This pattern of interpretation and revisionist history will be familiar in history and in those social disciplines characterized by narratives: stories revised continually as new archival evidence comes to light, making them intelligible to us, but never confirming as final the truths about what happened and why.

Introduction to the Literature

Laland and Brown, *Sense and Nonsense: Evolutionary Perspectives on Human Behavior* is an excellent introduction to alternative theories of cultural evolution, including both the nativist and the Darwinian cultural theories.

Among the most important recent works in the application of the theory of natural selection to explain cultural evolution is Boyd and Richerson's *Not by Genes Alone.* A paper by Richerson and Boyd is reprinted in Steel and Guala, which also includes articles by Dawkins and Sperber for and against the existence of memes—units of cultural inheritance and reproduction.

One work advancing a Darwinian theory of cultural evolution that makes hardly any claims about genetic hardwiring is Kim Sterelny's prize-winning alternative, *Thought in a Hostile World.*

CHAPTER THIRTEEN

Research Ethics in Social Inquiry

What is the connection between social science and ethics? This question is explored in the next two chapters. First we ask whether there are ethical constraints on social science inquiry. To deal with this matter we need to sketch relevant ethical theories. The three most prominent ones are briefly expounded.

ETHICS AND THE SOCIAL SCIENCE

In a poem titled "Under Which Lyre," W. H. Auden once enjoined his readers thus: "Thou shalt not sit with statisticians nor commit a social science." The injunction has been quoted in the past to suggest that there is something morally troubling about social science. In this chapter we examine some of the reasons for this suspicion. Like many other issues in this book, the suspicion turns on the divide between interpretation and prediction.

Right from the start there have been qualms about the consequences of either attempting or succeeding in the development of a predictively promising science of human behavior. For attempting may involve treating people in morally unacceptable ways, and succeeding may enhance our powers so to treat them. Indeed, predicting behavior may go further and encourage us no longer to view people as agents, responsible autonomous subjects of moral concern. Social science, some hold, is therefore inescapably dehumanizing.

There is also the prospect, which others take seriously, of morally dangerous knowledge. There are questions about humans that some say we should not take up at all, for answering them correctly can do no good and can only harm individuals and groups, even if it does not dehumanize them.

Social science cannot escape from moral issues in the way that many believe natural science can. For its subject is humankind. It provides the knowledge that guides human affairs, both individual and collective. More important, if there is moral knowledge, then we should expect social science to help provide it.

An atomic scientist might be excused from facing questions of what physics should be applied to—nuclear bombs or nuclear medicine—on the grounds that questions of application are not raised in physics. Can economists or political scientists avail themselves of the same neutrality? Or are moral prescriptions, prohibitions, and permissions built into their aims, theories, and methods? Even if a moral dimension is not built into theory, does the social scientist have a special responsibility to provide substantive moral guidance, a responsibility resulting from having a specialist's knowledge of human affairs?

Should we judge the acceptability of explanations and predictions in social science on moral standards and, if so, on which ones? These are questions social scientists answer either explicitly or implicitly in their work. But they are pretty clearly philosophical questions, and on balance, it would be best if the answers provided were both explicit and well informed.

MORAL PROBLEMS OF CONTROLLED RESEARCH

Assume that the goal of social science is predictively improvable explanatory theories of human behavior and that the methods appropriate to attaining this goal are roughly those common in natural science: hypothesis formation and the collection of observations and then the construction of experiments needed to test and improve the hypothesis. That brings us face to face with a moral problem that the physical sciences do not face and that the biological sciences face only on certain controversial assumptions. The problem is the conflict between the methods of empirical science and the rights of individual agents. In the biological sciences, the problem exists only on the controversial assumption that we need to consider the welfare or even the rights of nonhuman animals. Many proponents of animal welfare are struggling to restrict experimentation on animals. They argue against such experiments no matter how great the benefits medicine, agriculture, or cosmetology may secure from them for humans. Though most people care about the welfare of some animals (especially their pets), they reject the notion that animals have rights that enjoin us from raising them for food and killing them, let alone experimenting on them. But no one (publicly) defends the idea that people morally can be treated the same way as animals. And that makes for a large obstacle to empirical methods in social science. To see how large an obstacle, let's consider some real examples.

Perhaps the experiment that combines the most startling results with most serious ethical violations in social science is the famous "Milgram experiment." Stanley Milgram, a social psychologist, wished to examine the degree to which people would obey authority, even when their actions in obedience to authority would lead to what they believed were painful and even fatal results for others. Subjects were asked to help in an experiment on the effects of painful stimuli on learning tasks. They were to control a device that was attached to the experimenter's confederate; they were falsely told that the device could administer electrical shocks of varying strengths. The pretend subject would fail to accomplish the learning task, and the experimenter would order the real subject to administer a "shock." The pretend subject would emit an appropriate sound of discomfort. Successive failures would lead to instructions to increase the voltage and thus to louder and louder expressions of discomfort.

Milgram found that most people were willing to administer shocks they thought to be nearly lethal in spite of complaints, expressions of pain, and indeed feigned loss of consciousness by the mock subjects. People would do so, provided they were encouraged, authorized, or ordered to do so by an experimenter willing to take responsibility. After the experiment, each of the real subjects was debriefed in order to show that no harm had really come to the pretend subjects. However, Milgram reported that some harm had come to the real subjects as a result of this experiment. Some showed extreme nervous tension during the experiments, and some subsequently suffered from acute mental disturbance, despite the debriefing.

Two obvious questions are raised by this experiment: Was it worth it—do the benefits of knowing that ordinary people can behave like this outweigh the costs to subjects? Is it permissible to deceive subjects in order to acquire such knowledge? Note that without deception the experiment would not have been possible at all.

The conditions required by more passive observations lead to similar problems. Laud Humphrey's 1970 study of homosexuality reflects these problems. In order to study homosexual behavior and its social correlates, Humphrey began by volunteering to watch for police at the entrance to public facilities used for homosexual encounters. By recording the automobile license numbers of participants in these encounters, he obtained their addresses and, subsequently, posing as a public health survey taker, interviewed them. Humphrey's conclusions were that homosexuals in general led conventional lives, that a large number were married with children, and that they were no danger to the society as a whole. Indeed, he concluded that only ignorance of these facts and repression of homosexual conduct posed a threat to individuals and society at large.

Now, the consequences of Humphrey's research are widely viewed as beneficial. But the methods he was forced to employ, both by the nature of the phenomenon he studied and by the need to ensure reliable observation, bring up several serious moral questions. In addition to the several deceptions he had to practice and his violation of his subjects' privacy, the risks to them could have been very great. At the time that Humphrey was conducting his research, great harm could have been done to his subjects if his data, including names and addresses, had fallen into the hands of the unscrupulous. Do the benefits of enlightening research like this outweigh its potential risks? Humphrey, of course, knew the potential damage his data could do and acted to minimize the risks. But can we sanction such research even when the risks are quite low? As an aside, notice that our social science will have to be reliable enough in its predictions to enable us to estimate these risks.

Sometimes observation and experiment have morally questionable effects beyond the immediate subjects. In the 1950s, social scientists secretly recorded the deliberations of six different juries in Wichita, Kansas, in order to provide empirical data to test certain assumptions about the US legal system. Their aim was to improve the system of justice by providing information that might help better understand it. But these social scientists may have put the system of justice at risk. For when the study became known, the confidentiality of jury trials everywhere was undermined and, with it, a fundamental provision of the Bill of Rights to the Constitution was compromised. Moreover, though no harm came to jurors or defendants in the study, their rights were certainly violated.

To explore the moral problems these cases raise, we need some tools, principles that reflect our moral convictions and theories that explain, justify, and enable us to apply these convictions. Similar problems in biomedicine have led to the identification of certain moral principles to guide research. It may be convenient to consider them and their application to the social sciences.

One principle illustrated by our cases is that individuals have some rights that cannot be abridged and that they must be allowed to make their own decisions autonomously. Autonomy can be reduced by failure to inform persons of the circumstances in which they act and by failure to secure their well-informed consent to research in which they will be subjects. Being well informed is not enough, however. Autonomy is reduced when informed consent is secured under circumstances of implied or explicit coercion. Thus, it is sometimes held that even telling a prison inmate about all the possible effects of an experiment for which she may volunteer cannot be respectful of her autonomy. For, as a convict faced with the inducement of a reduced sentence for good behavior, her consent may not be fully autonomous.

In medicine, this principle seems less restrictive than in social science, though it certainly reduces the supply of experimental subjects. For except in the case of placebo experiments, deception does not seem essential to an experiment's success, once a subject's agreement to the experiment is secured. But because of the reflexive nature of social science, a subject's behavior is likely to be influenced by learning what hypotheses the investigators are studying and what methods they are employing. Therefore, a blanket prohibition against deception, on the grounds that it violates autonomy, will severely restrict social science. This means that if we employ deception and if we need to violate the subject's right to autonomy, we also require a moral justification.

Another principle important in medicine is the injunction not to inflict harm on subjects and to benefit them to the fullest extent possible. The first part of this principle, nonmaleficence—doing no harm—takes precedence over the second, beneficence, making positive improvements. In medicine this principle restricts the scope of experiments greatly. It requires empirical studies to proceed slowly and with the greatest caution. Before it can be determined that a new drug can help anyone, it must be ascertained that the drug will do no positive harm. And in medicine when a procedure provides both harms and benefits, there is a presumption against it because of the priority of nonmaleficence over beneficence. Even where benefits are held to outweigh harms, the greatest difficulty lies in how we measure them in order to make comparisons.

In the social sciences, applying the principle of the preponderance of benefits over harms to the treatment of experimental subjects, who presumably have given informed consent, is even more difficult. Employing this principle not only is more difficult but also exposes moral choice based on cost-benefit analysis to a frustrating vicious circle. In order to expand knowledge in the social sciences, we may have to undertake experiments. For these experiments to be morally permissible on the present principle, we need to know what harms and benefits we can expect our experiments to produce and perhaps also to weigh them against each other. But determining harms and benefits requires the very sort of social scientific knowledge that the experiments are designed to produce. What is more, economic theory tells us that harms and benefits cannot be given cardinal weightings, nor can they be summed between different people, because of the incoherence of cardinal or interpersonal comparisons. Therefore, applying this principle will be impossible on substantive theoretical grounds as well as on logical and methodological grounds.

There is another moral principle generally more relevant to research in social science than in biomedical contexts. It is the principle of fairness, equality, and justice. It requires equal treatment of individuals, equality of opportunity, and equal access to potential benefits and advantages. For example, large-scale studies of the effects of social policies, like a negative income tax or free day-care centers, involve randomly selecting some individuals to receive the benefit and others not to, followed by comparison of the effects of this differential treatment. But subjects relegated to the control group are deprived of the experimental group's benefit. They may complain that they have been treated unfairly, their right to equal treatment having been violated.

As with other ethical problems, there are ways around this issue, but they involve forgoing experimental designs in favor of "quasi-experiments." Such studies require the imaginative use of data already available, often data collected by government agencies for other purposes. Intelligent observation of phenomena that social scientists have not themselves arranged shifts the burden of injustice or unfairness to whoever is responsible for the unequal treatment to be studied. Moreover, such unobtrusive measures also avoid some problems of autonomy, such as informed consent and the nonmaleficence/beneficence calculation, for the subjects studied at any rate. But those measures do not circumvent all the moral problems of social science research. There remain threats to privacy and confidentiality and the balancing of harms and risks to social institutions and public confidence in them. But clearly, such nonexperimental alternatives are a second-best solution in a domain where even the best approach often provides little insight itself.

The moral principles that restrict inquiry in social science raise two broad philosophical problems. The first is a double question. What is their own justification? And why should we obey them? For many, it is enough simply that they are widely recognized and historically venerable principles to which almost everyone gives assent. Moral rules lie at the foundations of civilized society and therefore are beyond question. Others find the grounds for adherence to these principles of conduct in the teachings of various religions. Philosophers are rarely satisfied with either answer to the question of how moral principles are justified. The first answer—that these moral rules are obviously true—is undermined by the fact that in many societies some of them are not accepted. And where moral rules are accepted, mere acceptance, even when universal, is no sure mark of truth or justifiability.

The second answer, that these principles are God's command, has little force for the scientist. More important, whether respecting autonomy is morally right because it is enjoined by the Lord or whether it is enjoined by the Lord because it is morally right is left an open question. Presumably it is the latter, but then we need to ask what it is about this principle that recommends itself to the Lord as the morally right one. And that is our original question all over again: What justifies these principles as the morally right ones?

One answer we must exclude is that we should obey these principles because they are written into the statutes and common law of our legal systems. Like others, social scientists are legally responsible for their actions. They may be held liable by individuals or by public prosecutors for harms they knowingly or even unwittingly inflict on subjects and nonsubjects. And governments have adopted laws inhibiting certain methods and subjects of research because of their morally disturbing ramifications. The Wichita jury study, for instance, led to legislation that explicitly prohibits the recording of federal jury deliberations. And in recent years, the US government's chief sponsors of social scientific research (the National Institutes of Health, the National Science Foundation, the Department of Education and the Department of Health and Human Services) have promulgated regulations to govern inquiry it supports. The regulations reflect most of the principles described above and require review boards to supervise research on human subjects in order to prevent ethical abuses.

But the authority of such agencies does not solve the problem of what the moral justification of such principles might be. At most it provides a prudential justification for them. Social scientists who fail to abide by them may be at risk of criminal or civil prosecution. In fact, governmental regulation raises the same moral problem all over again. For we need to examine whether and why governments have the moral justification to enforce these principles in the conduct of research.

Another problem these principles raise is how to adjudicate conflicts between them. Sometimes, both within and beyond experimentation and observation, the principles pull us in different directions: preserving autonomy may prevent us from providing "benefits" to those unwilling to accept them, mistakenly or not. Sometimes we must forgo potentially great benefits to society, if we are to preserve individual rights. Sometimes we must violate those rights to secure such general benefits or to avoid great harm to large numbers of people. How do we decide these cases of moral conflict?

Both the problem of justification and the problem of moral conflict call for a general moral theory that will underwrite these principles and establish a hierarchy among them when they come into conflict. The trouble is that there seems little agreement on such a theory among philosophers and others. And the leading candidate theories simply embody the conflicting dictates of the moral principles we have summarized. What is more, the leading candidates among moral theories reflect the two divergent trends in methodology of social science that we have examined in many of the previous chapters. Thus, taking sides on questions of scientific method in the social sciences may commit us to taking sides on fundamental matters of moral philosophy.

The two leading accounts of the foundations of moral judgment and the grounding of moral principles are utilitarianism, attributed to British empiricist philosopher and economist John Stuart Mill, and a theory of duties and rights developed by German philosopher Immanuel Kant. The latter is often called a *deontological* theory, from the Greek for "theory or knowledge of what is binding or obligatory."

MILL: NATURALISM AND UTILITARIANISM

Utilitarianism is a moral theory originally developed by one of the most important and influential economists of the nineteenth century, John Stuart Mill. In addition to his work in economics, Mill made great contributions to the philosophy of science and especially of the social sciences. But he is most well known for many influential works in political and moral philosophy, which reflect the influence of the social sciences, especially theories in psychology and economics. Mill's moral theory continues to be taught side by side with the theory of rational choice by welfare economists.

Utilitarianism holds broadly that actions are to be assessed as morally acceptable or not on the basis of their consequences for the happiness, satisfaction, welfare, or *utility* of those affected by them. Thus, any moral principle must be assessed in terms of the consequences of its adoption for all affected parties. To see its immediate application to research ethics, consider how it decides whether we ought to respect the autonomy of any individual or treat this person strictly on the basis of nonmalfeasance and benevolence. To do so, utilitarianism insists, we need to calculate the effects on all individuals of such treatment of this person. If the costs to everyone of such treatment outweigh the benefits to everyone, then, at least sometimes, we must forgo the principles and must violate rights in order to attain a more generally desirable outcome.

It sometimes seems that a violation of individual rights will result in a greater benefit for a larger number of people. More often, the reverse seems true. Either way, if we can weigh the benefits, utilitarianism gives us a means of resolving conflicts between moral principles pulling in different directions. Furthermore, some utilitarian philosophers have tried to show that violating rights never really results in greater benefits, so there is no conflict between benefits and rights when utilitarian principles are properly employed.

Utilitarianism is pretty clearly a bedfellow of the naturalistic approach to social sciences, and its weaknesses reflect problems in the development of a predictively improvable science of human behavior. Utilitarians initially advanced their doctrine on the assumption that welfare or satisfaction could be measured cardinally and compared among people and that people are utility maximizers—rational economic agents. They further supposed that we can foresee (that is, predict) the consequences of certain actions, rules, or policies for people's welfare. Originally, the principle was stated more explicitly as requiring us to adopt the actions, institutions, and practices that maximized average utility, total utility, or the utility of the least advantaged in a society.

With the eclipse of cardinal and interpersonal utility, the principle had to be considerably weakened. In welfare economics, utilitarian prescriptions that we maximize the greatest happiness became the requirement that we adopt a policy if it can be shown to increase at least one agent's utility without decreasing the utility of any other agent (called the Pareto principle, after its first expositor, Vilfredo Pareto). That means, of course, that we cannot deprive a millionaire of $100 and be sure of a net increase in total utility when we give the $100 to a pauper because we cannot be sure that the loss in utility to the former will be smaller than the gain to the latter (see Chapter 6). Nor will we be able accurately to balance the costs of violating someone's privacy or deceiving someone in a social experiment against the benefits of the knowledge produced for all affected individuals.

As we have seen, this restriction has led to serious charges leveled against modern welfare economics by critical theorists, among others (see Chapter 8, "Critical Theory"). Even if the charges are unwarranted, the possibility that a utilitarian theory will solve our problem of adjudicating moral conflicts seems slim.

But the core of utilitarianism seems right to many people because it focuses moral decisions on consequences for all affected parties. Though we cannot measure welfare very well, we all believe there is such a thing as individual welfare, and it seems right that policies should be judged in the light of how they affect everyone's welfare to the extent we can measure it. Thus, many who will not accept utilitarianism are, nevertheless, *consequentialists*. They ground their moral decisions on the consequences of policies for affected parties. But consequentialism—deciding on policy by looking to consequences—makes very strong demands on social science.

To determine the consequences of a policy, we need to know its effects with considerable accuracy. We need to know how the policy will interact with other social and natural processes as they influence people's welfare. Even if all we need is knowledge of whether more people will be benefited by a policy than harmed by it, the demands on social scientific knowledge will be very great. And the demands are predictive ones. They can be met only by uncovering laws and generalizations that, when combined with descriptions of the policies we adopt, are what will enable us to derive projections about the future. So consequentialism requires a predictively improvable science of individual behavior and its aggregation. The need for predictive power will be true for consequentialist decisions about whether we should undertake experiments to further social science. It will be equally true for consequentialism as a basis for public and private policy of many different kinds.

Conversely, limitations on our ability to provide such predictive knowledge will severely limit a consequentialist approach to many vexing social problems. For if we have no confidence about the effects of a policy, consequentialist theories provide no recommendation at all. If prediction beyond the powers of folk psychology is impossible, then consequentialist justification for many governmental policies is a dead letter.

Consequentialism is so closely tied to a naturalistic approach to social science that it may turn out to seriously undermine some of the moral principles described above. Suppose that the unsuitability of notions like belief and desire, and especially action, to a predictively powerful science leads us to surrender them as accurate descriptions of human behavior and its consequences. Behaviorists as well as some contemporary neuroscientists certainly endorsed this view. But surrendering these concepts as unscientific also requires us to give up other concepts built out of them. That means giving up the intelligibility of concepts like autonomy, freedom, dignity, responsibility, informed consent, and privacy. *Autonomy*, for instance, means doing what you really would want to do if you had all the facts—that is, the relevant true beliefs.

If these concepts turn out to have no explanatory function with respect to human behavior and its causes, then a consequentialist approach to social policy can safely neglect them. Thus, for example, the behaviorist B. F. Skinner argued that our system of criminal justice, educational institutions, and political processes are inefficient in meeting our goals because they are based on theories of human behavior that embody these vacuous notions. In effect, Skinner held that these notions will ultimately suffer the same fate as, say, the concept of witchcraft or demonic possession.

In the fifteenth and sixteenth centuries, the behavior of the mentally disturbed was explained by appealing to such notions, and the treatment meted out to such persons reflected the theory that their behavior was caused by the devil. Doubtless, contemporary treatment of the mentally ill is far more humane, which in part reflects our decision to forgo explanations that appeal to the concept of witchcraft. One way we put forth this point is by saying there are no such things as witches. Similarly, Skinner held, it may turn out that our treatment of criminals, for example, would be more humane and more effective in rehabilitation if we stop regarding them as autonomous agents, "responsible for their crimes," and instead view them as having been exposed to environments that brought about their crimes and that must be changed in order to elicit socially more acceptable behavior.

If the social sciences must abjure notions of autonomy, responsibility, and free will in order to provide the kind of theory consequentialism needs, then they can hardly be expected to justify moral and legal principles embodying these very notions. In fact, many moral conflicts will turn out to rest on our false beliefs about the causes of human behavior. Thus, the advance of science would dissolve them without recourse to fundamental moral theory.

KANT: INTERPRETATION AND DEONTOLOGY

The most widely endorsed objection to consequentialism leads us directly to a deontological moral philosophy, one that is as closely related to non-naturalistic approaches to human behavior as consequentialism is connected to naturalistic ones. Suppose someone offered you the opportunity to enrich the lives of a hundred or a thousand or a million persons in return for the opportunity to kill by the most painless means possible a randomly chosen child. Though many of us might accept such an offer, most people would agree that doing so would be morally wrong. Why? Because we would be violating the rights of an entirely innocent person. It seems a deep-seated conviction that at least some rights cannot be trampled on, no matter the beneficial consequences. Indeed, our moral conscience holds that some rights cannot even be given away or sold willingly by their bearer. Recall the Declaration of Independence: "All men . . . are endowed . . . with certain unalienable rights . . . among these are life, liberty and the pursuit of happiness." And these are not mere words. They reflect a moral theory fundamentally opposed to consequentialism. This is the theory that human beings have certain special duties and rights and that these rights and duties always take precedence over policies that would violate these rights or excuse these duties for some greater good for all.

Where do these rights and duties come from? According to Kant, they emerge from pure reason, untainted by experience. Reflection alone leads us to see the truth of certain a priori moral principles, that is, principles that we can know for certain without experience. Because they are necessary truths, all possible experience must be compatible with them; hence, no particular experience can justify them. The principle Kant identified as paramount he called the *categorical imperative*. This rule enjoins all rational agents to endorse for themselves only principles that could be endorsed by all moral agents. The only such principles are those that require moral agents always to be treated as ends and never as means to some purportedly greater end. Followers of Kant hold that treating people as "ends in themselves" entails not allowing the consequences for some or even many to influence our treatment of others, no matter how few. Each individual has certain rights, and we have certain duties to respect those rights, come what may. That is what is involved in individual autonomy. It is what grounds the moral prohibition against deception, the violation of privacy, and unequal treatment. Autonomy trumps the achievement of good outcomes and the avoidance of bad ones. Thus it subordinates the principle of nonmaleficence and beneficence to autonomy, and thus resolves the conflict between them in the opposite direction from the resolution effected by utilitarianism.

What is more, the focus on autonomy makes few demands on improvable scientific knowledge of human behavior, because such knowledge cannot influence our moral judgments about what is permissible and what is not. Kant's theory would positively prohibit steps to increase that knowledge if such steps require violations of human rights. Indeed, it may encourage a conception of human behavior that makes predictively improvable social theory unnecessary and the attempt to formulate it misguided. It makes that attempt misguided in two ways.

First, the conception of rights and duties that deontological moral theories require makes the notions of belief, desire, and action fundamental to our conception of ourselves. Without them, morality has no foundation in this view. For rights are rights to perform actions, and duties are duties to do so as well, and actions are necessarily the outcome of beliefs and desires.

According to Kant, one of the features of an ethics that makes rights and duties paramount and subordinates consequences is that moral assessment must focus on motives for actions, instead of their consequences. A beneficial act done for the wrong motives will have no moral value. A harmful act done with the best of intentions and the right motives may be morally praiseworthy. The essence of moral rightness for Kant was action motivated by a desire to do one's duties and carry out one's obligations. And, according to Kant, the relations between our beliefs about our duties and our desire to fulfill them and the actions these beliefs and desires explain cannot be understood causally at all. For to treat the relation as causal would rob human action of its moral dimension altogether. The reason, Kant held, is that causality implies determinism and thus the absence of free will. But free will is a prerequisite for moral responsibility. It is pretty clear how much Kantian moral theory leans on a social science that takes human action and intentionality seriously, and rejects causal determinism.

Second, Kant claims that we can have a priori knowledge—that is, perfect certainty about the truth of some statements without recourse to experience. This possibility fosters the conviction that we might acquire further knowledge simply by pure reason, by reflection unaided by observation. In particular, pure reason may also be the source of the conception of intelligibility that animates noncausal approaches to the explanation of human behavior. Kant defended the categorical imperative on the grounds of its intelligibility to pure reason. He rejected consequentialist theories in part because their truth could not be warranted by intelligibility alone. The same criterion of intelligibility might be enough to certify the explanatory power of a principle like [L] or some more profound insight into human action, just because it makes behavior intelligible.

It should be noted that Kant identified a moral principle—the categorical imperative—as justified because it was the only one that could recommend itself to pure reason. He did not suppose that particular moral decisions could be made by pure reason. Such decisions must be taken by applying a priori moral principles to details about particular cases, details that only experience can provide. Similarly, an explanatory principle is itself justified a priori, not as a causal law, but as a principle constitutive of action. But it cannot be applied to particular circumstances except by adding facts available only through experience. The point is not that social science is a priori but that the power of its explanatory strategy is based, not on experience, but on pure reason.

It is worth noting the affinities of Kant's theory of knowledge and his moral theory to Habermas's critical theory: like Kant, Habermas sought a realm of knowledge beyond and independent of naturalistic empirical understanding, and he sought from it guidance about how to use empirical knowledge of human affairs to attain outcomes certified by reason as the morally right ones, ones in accord with human dignity and reason. Though critical theory does not endorse Kant's explicit categorical imperative, it shares with Kant a commitment to the possibility of a priori knowledge that controls empirical science.

Little of Kant's complex argumentation for a deontological approach to morality has been acceptable to most philosophers. Therefore, other foundations have been sought for principles that give autonomy priority over consequences in deciding moral conflicts. One tradition, which stretches back well before Kant, accords individuals *natural* rights and, concomitantly, natural duties to respect those rights. A natural right is one that each individual has just by virtue of being human. It is part of what makes one a person, in contrast to a merely very complicated system composed of organic material. But any serious theory of natural rights must explain exactly what rights we have essentially and what it is about people that confers upon them such natural rights. Presumably, natural rights will have something to do with the fact that we are sapient and sentient creatures, that we have thoughts and feelings, intentional states, and minds, which other creatures lack. (Notice that those who also accord such states to animals will find it more reasonable to also attribute rights to them.)

Like Kant's moral theory, a natural rights view seems to rest on a nonnaturalistic conception of human beings, one that exempts them from a thoroughgoing scientific treatment. Despite the confusing similarity of terminology, a naturalistic theory of human thought and behavior has no room for distinctive natural rights. It assimilates humans to the rest of nature and denies any essential distinctiveness of people. Naturalism aims at a causal account of their behavior and their minds that shows humans to be no different in kind from other, simpler systems to which we do not accord rights.

THE SOCIAL CONTRACT AND RAWLS'S THEORY OF JUSTICE

There is yet a third theoretical tradition in ethics, one with historical roots among the founders of empirical social science and with great contemporary influence especially among students. This approach attempts to derive and justify moral principles by showing that rational agents, endowed with certain interests, would agree among themselves to adopt certain moral principles, including a set of strongly binding rights. They would do so because it would appeal to them as rationally self-interested agents as part of a contract among themselves for the organization of society. As noted, social contract theories were advanced by many of the fathers of modern social science, philosophers like Hobbes, Locke, Rousseau, and others. Their object was to ground the moral principles we recognize by showing that any rational individual would agree to them rather than live in a condition of anarchy, without the advantages of moral rules.

If, as rational choice theory claims, all individuals seek their own advantage, moral principles will be endorsed and obeyed by all individuals just in case the benefits to each individual of adopting these principles outweigh their costs. Social contract theorists of the seventeenth and eighteenth centuries disagreed among themselves about which particular moral principles rational individuals would endorse. In the nineteenth century this approach to ethics fell out of favor. But in the twentieth century the general strategy of justifying and identifying moral principles that social contract theory envisions began to exercise a renewed fascination as rational choice theory and evolutionary game theory developed.

Perhaps the most influential of these social contract theories in the social sciences is due to John Rawls. The theory is not difficult to sketch, its influence is enormous, and it provides answers to our questions about the morality of social scientific research that contrast with those of Mill's utilitarianism and Kantian deontological theories. Rawls's theory, like utilitarianism, also shows the influence of economics, especially rational choice theory, and of significant assumptions about choice made by cognitive social psychologists.

Reasons for the renewed interest in social contract, or as they are sometimes called, contractarian theories stem from some of the difficulties of utilitarianism and the limitations economists imposed on the theory of rational choice, which utilitarianism spawned. If interpersonal comparisons are impossible, utilitarianism cannot choose among many different moral rules or social institutions, all of which pass the Pareto test. What is more, utilitarianism cannot provide moral decisions in the absence of reliable information about consequences that we have no way of providing. Some economists and philosophers have hoped that social institutions and moral rules could be derived from the examination of rational strategies or from hypothetical bargaining problems faced by perfectly rational economic agents.

The chief trouble that these theories face is that moral rules encourage free riding and the rational desire to avoid being made a sucker. In the context of the prisoner's dilemma, the "moral" thing to do is to cooperate with the other player, especially not to break one's word if a promise or a contract has been entered into not to confess. But confession, as we have seen, is just what rationality enjoins. Work in this area continues among economists and philosophers hoping to provide a contractarian justification for moral principles. Once such a justification is found, the economists and philosophers will have to face the problems of relating the unrealistic and idealized assumptions that social contract theory shares with economic modeling to the decision procedure of real people.

Strictly speaking, Rawls's theory is, as his famous book is titled, *A Theory of Justice.* Thus it is explicitly a political philosophy, and not just a moral theory. But, insofar as utilitarianism and Kantian theories are applied to the design and operation of social and political institutions, this difference is not particularly important. Thus, rules for the treatment of human subjects in social research are matters of justice as much as welfare.

Rawls's theory begins with a bargaining problem. Each individual is a participant in a hypothetical negotiation with all other individuals bargaining to the set of rules that will govern society. In order to ensure objectivity in the bargaining, each bargainer must operate behind a "veil of ignorance" about his or her special abilities, skills, tastes, gender, intelligence, cultural values, religion beliefs or practices, and economic endowments once the rules begin to be enforced. However, each party has information about the general facts regarding the society, and knowledge of social science relevant to assessing various proposed rules, practices, and institutions. Thus, all recognize that private property and taxation will be necessary for all reasonable arrangements. Behind this veil of ignorance the only other fact people know about themselves is that they are sufficiently risk averse that each person will choose the package of rules and institutions that will maximize their position in society, should that person turn out to be the worst off once the veil is lifted. Thus Rawls makes a strong, substantive, empirical psychological assumption that rational agents are *maximin* strategists, maximizing our minimum possible outcomes. By contrast, other philosophers and economists claim that people either should be risk neutral or in fact willing to accept the chance of a well below average outcome in return for the chance of a well above average outcome. Rawls's maximin assumption is crucial for his argument about what justice requires. But because it is a factual one open to disagreement, the foundation of his theory of justice is no stronger than this single assumption.

Given that all rational agents are maximin strategists operating behind a veil of ignorance, Rawls argues that we will bargain to an outcome that has the following specific features:

1. The maximum amount of fundamental basic political rights and liberty each person can have, consistent with all others having the same set of basic rights and liberty.

These Rawls calls *primary goods*, which no one can do without, especially if one finds oneself worst off in society after the veil is lifted. These are so important that justice precludes their being traded away for any amount of improvements in general welfare. Thus on certain rights, Rawls sides with advocates of rights and duties against advocates of utility or welfare maximization.

2. Equality of opportunity for all places in the society.
3. "The difference principle": inequalities in the society will be accepted provided that they enhance the welfare of the least advantaged representative persons in the society.

The third condition is a weakly egalitarian one, allowing inequalities in income and wealth provided that institutions that permit them are to the advantage of the worst off. For example, paying higher than average salaries to physicians will attract the most qualified to this difficult job, which is to everyone's advantage, including the worst off. Here Rawls sides with utilitarians against those who would treat, for example, economic rights as more important than considerations of welfare. The difference principle, along with other incentive-sensitive features of Rawls's conception of just institutions, reflect insights about the efficiency of the free market that go back through economics to Adam Smith. They also reflect his appreciation of the fact that the market and other social institutions have (latent) functions for members of societies that they do not themselves recognize. Moral theories like Rawls's that rest on the existence of such institutions need, of course, like factual theories in social science, an account of how such institutions can have arisen and how they can be maintained, even though no one designed them and no one is acting to sustain them.

Widespread interest in Rawls's theory, and indeed support for it by contrast to utilitarianism, is due in some measure to its rejection of the strictures and limitations that rational choice theory, and especially ordinal utility theory, place on moral judgments about fair distributions. Recall the critical theorist's analysis of the ideological service that economic theory plays to capitalist inequalities: by making impossible interpersonal utility comparisons, it "shows" there is no way to tell whether giving a homeless man $100 makes him better off than giving it to a millionaire. Thus economics can't justify redistribution to maximize welfare, and the wealthy have an argument that redistribution can't be proved to make the whole population in general or on average better off. Rawls's theory enables a critic of this claim to provide an alternative foundation for making such judgments.

In the present context, Rawls's theory will allow scientific research and expenditures on it to the extent that any inequalities in outcomes it produces have advantages for the least well off in a society, and so it will underwrite a good deal of such research. The difference principle allows support for social research on human subjects so long as it does not violate the rights that constitute primary goods and confers advantages on the least advantaged members of society. Thus it may prohibit studies such as Laud Humphrey's or the Wichita jury study, if they violate the first principle of maximum liberties—the primary goods—for all.

The role of the substantive psychological maximin principle, and the demand that distributions constitute Pareto optima, in which the least advantaged are given a veto, show the extent to which Rawls's theory reflects agendas of research in the social sciences. Indeed it may be hostage to developments in social science owing to these commitments. Rawls's theory and the book in which it was expounded make contact with many social sciences and their results, not just the maximin and Pareto principles. Accordingly, it is subject to expansion and perhaps also revision, in its derivation of the principles of justice from general facts about rational agents operating behind a veil of ignorance.

CONCLUSION

Consequentialist moral theories, Rawlsian theories, and deontological ones are advanced to underwrite moral convictions. The theories are supposed also to order, or prioritize, the application of moral convictions, not just to experimentation in social science and elsewhere, but to all areas of social and individual life. But three such irreconcilable theories with differing ethical implications leave us with larger problems than the ones we hoped the theories would solve. Instead of justifying and ordering relatively concrete moral convictions, we are now faced with settling fundamental disputes in moral philosophy about the correct abstract, general ethical theory. Moreover, it appears that the two views about the nature of social science that we have examined in this book are closely associated with these divergent moral theories. In choosing between different methods in social science, we may also be making moral choices or at least choices about the adequacy of these moral theories.

We may make such choices, but perhaps we need not. For despite the sympathy that each of our philosophies of social science may have for a different moral commitment, we have not shown that any of the three entail an ethical theory. At most we have seen affinities. Consequentialism's relevance to real moral decisions requires a predictively improvable social science. Deontology's commitment to rights and right motives for action presupposes the explanatory powers of intentional states, their predictive weakness notwithstanding. Rawls's theory turns on a qualified commitment to rational choice. But, it is often alleged, none of these moral theories is logically incompatible with any purely factual approach to human behavior. So advances in social science will neither solve nor enable us to avoid moral conflicts. It is on this argument that the next chapter focuses.

Introduction to the Literature

T. Beauchamp, R. Faden, J. Wallace, and L. Walters, eds., *Ethical Issues in Social Science Research*, is an excellent introduction to the issues and contains many important papers. S. Milgram, *Obedience to Authority*, and L.

Humphrey, *Tearoom Trade*, report examples of research that raise serious moral questions.

An excellent introduction to ethics and moral theories is W. K. Frankena, *Ethics*. Very different views are advanced in G. Harman, *The Nature of Morality*, and J. L. Mackie, *Ethics: Inventing Right and Wrong*. J. S. Mill, *Utilitarianism*, and I. Kant, *Foundations of the Metaphysics of Morals*, present the two moral theories treated in this chapter. Kant's work is extremely difficult for the student to understand. The notion of natural rights stems from the thought of John Locke and is firmly entrenched in modern legal and political doctrine. A radical development of it is to be found in R. Nozick, *Anarchy, State and Utopia*. More influential but less easy to classify is J. Rawls, *A Theory of Justice*.

A. Buchanan, *Ethics, Efficiency and the Market*, traces the connection between moral theory, especially utilitarianism, and economics. B. F. Skinner's claims about the consequences for moral philosophy of behaviorism are advanced in *Beyond Freedom and Dignity*. A general treatment of ethical issues surrounding behaviorism is E. Erwin, *Behavior Therapy: Scientific, Philosophical and Moral Foundations*.

CHAPTER FOURTEEN

Facts and Values in the Human Sciences

It has often been claimed that unlike theories and explanations in natural science, the ones that social scientists advance have strong moral, normative, evaluative presuppositions, commitments, or consequences. The avoidability or unavoidability of values in the description and explanation of social facts is a question that has long vexed all the social sciences. Many empirical students of human affairs have insisted that their findings and theories are value free. Others have denied the very possibility of neutrality. In recent decades this dispute has been joined by those who have argued that much social science reflects values inimical to women and other minorities, and to their contributions to social knowledge as well.

FACTS AND VALUES IN THE HUMAN SCIENCES

According to some philosophers, the connection between ethics and social science is even closer than an affinity. Therefore, in social science we always take sides on moral questions. If that were so, it would be a crucial difference between natural and social science. In fact, some hold it to be the source of what has been described as the difference between natural science's relative progress and the alleged lack of progress in social science.

Much social science has been driven by the moral values and ethical imperatives of social scientists. Just choosing what explanatory or predictive problem to work on, what phenomena to understand, or whether to interest oneself in a particular social process is often the result of an initial evaluation that how things are done, or their outcome, is unacceptable, can be improved, is unjust, unfair, inequitable, needs to be changed. The same is, of course, true in natural science research. But the special issue for social science is whether the results of inquiry in the social sciences, the findings, theories, data, explanations, et cetera, are themselves neutral as between differing values and commitments. It is widely held that objectivity in natural science requires neutrality. The question raised by philosophers of social science and others is whether neutrality is possible in social science, and if not, what the ramifications are for objectivity in social science.

NORMATIVE VERSUS POSITIVE; PRESCRIPTION VERSUS DESCRIPTION

To understand this issue, we need to understand the difference between facts and evaluations, or description and prescription. Our task is complicated by the fact that many who argue that social science always takes sides on moral questions do so by denying this distinction. Therefore, whatever one says to introduce it is bound to be unacceptable to some parties to the debate.

A factual claim describes the way things are, while remaining neutral on the question of how they ought to be or whether they are good or bad or could be improved or worsened. A claim about what is the case is *value free*. A normative or evaluative or *value-laden* statement expresses values or evaluations of facts based on those values. Or it may both describe and evaluate. That is, it can express approval or disapproval, praise or blame, for the fact it also describes; it can reflect the suggestion that the fact ought not—or ought—to be the case. A simple example is the contrast between saying that Lincoln was killed and saying that he was murdered. The former states the facts but is neutral on whether his killing was a wrongful death. The latter reports the same fact but takes a stand on whether it ought to have happened. "Lincoln was murdered" presupposes some moral theory about what ought to be and what ought not to be the case, about whether some killings are morally permissible. Such statements reflect the ethical norms, often unexpressed, of the speaker.

One traditional view about science is that it is value free, or morally neutral. The theories, laws, experimental descriptions, explanations, and predictions of physics or chemistry seem quite independent of any ethical teaching. Of course, natural scientists make value judgments, and some of these will be informed by the scientists' specialized knowledge. But in so doing, they express views that follow, not from their scientific beliefs, but from those beliefs combined with their independent moral beliefs. For example, a physicist's opposition to a new weapon system may derive from his belief that it is unworkable or that it is too expensive. In each case, his opposition to the system will follow from these beliefs only if we add in evaluative premises. In this case the premise is that money should not be spent on physically unworkable systems or spent on systems that are not cost effective. Though such principles may be obvious, they rely on other, more basic moral claims: that it is morally wrong to waste scarce resources, for example.

Some philosophers have held that moral claims cannot be part of science because such claims do not constitute knowledge at all. Rather, moral judgments are expressions of emotion, taste, or subjective preference. One argument in favor of this view is the fact that people who seem to agree on a very wide range of factual questions may yet disagree about the most fundamental moral ones. Thus, two physicians may agree on all the facts about a particular prospective abortion, including all the physical, psychological, and social consequences for mother and fetus of having the abortion; yet they may still disagree about whether the abortion is morally permissible. A second consideration given in favor of this skeptical view of moral knowledge is the fact that moral teachings differ widely among cultures, subcultures, and ethnic groups. Since it seems ethnocentric to insist that some of these teachings are false, it has been concluded that none are true. But if there is no such thing as true and false when it comes to values, then ethical theories, no matter how firmly believed, cannot count as knowledge.

It follows from this skeptical view of ethics that if social science is to be knowledge, it ought to emulate the value freedom of natural science. Social scientists ought to be careful about how they express their findings and theories in order to ensure that value-laden descriptions don't contaminate them. They should be scrupulous about labeling any evaluative claims as such. Thus, like the physicist, a political scientist can oppose a weapons system as *undesirable* because it is politically destabilizing. Here the political scientist's specialized knowledge of the effects of such a system on international relations is crucial, but his opposition follows only from this knowledge plus a normative claim that destabilizing policies *ought* not be pursued.

Even if moral knowledge is possible, the persistent disagreements about it among those who share many other beliefs, suggest that acquiring such knowledge is difficult. Indeed, the tolerance that characterizes most Western societies reflects the belief that moral questions are difficult to answer with much unanimity. Moral certainty breeds paternalism, if not intolerance and, ultimately, totalitarianism. The avoidance of moral absolutism is one reason to favor *value neutrality* as a methodological principle for social science, even for those who do not demand that it emulate the features of natural science.

The conviction that social science should, like natural science, be value free is widespread among experimental psychologists, economists, and the more quantitative of social scientists. Economists especially insist on advancing what they call a *positive*, as opposed to a *normative*, science. They hold that their discipline cannot make substantive policy recommendations because such conclusions are normative. Economists can only trace out the consequences of various policies, actual and possible, leaving it to the politician to decide which should be implemented. In fact, the economists' self-imposed restriction of economic theory to ordinal utility and revealed preference, and of their welfare criterion to Pareto optimality, is most often justified by this commitment to a purely positive science.

Many exponents of positive social science provide a moral or, at any rate, a prudential argument in its favor. They hold that it is important to avoid a normative bias in social science, for that can destroy the objectivity crucial for informing social policy. Evidence slanted by personal values, conclusions shaded to advance individual preferences or theories reflecting implicit commitments, even the highest moral conceptions, all may destroy both public confidence in social science's objectivity and factual reliability. Such a bias may frustrate the very aims social science is called upon to guide. Sometimes it may be difficult to attain the sort of moral neutrality required. But the social scientist has an obligation at least to be explicit about the values held, again because they may color judgments and impair objectivity. When that happens, the information a social scientist offers to inform policy will reflect biases and impede the attainment of social goals chosen by the society through democratic procedures. Just as few would wish to have someone else's values imposed on them, so the social scientist has no right to impose moral standards on society.

But remaining value free is, according to many, far more difficult for social science than for natural science. According to others, it is flatly impossible; still others hold it to be undesirable. The argument that moral neutrality is impossible for social science often goes together with the argument that such neutrality would be itself morally undesirable. Together, the arguments purport that what would have to be done to free social science from value commitments would result in something no one would recognize as a science of human action that explains events by uncovering their significance. At best the result would be a powerful tool for social control; at worst it would just be an empty exercise in "physics envy." Both of these outcomes are held to be morally repugnant. They threaten our view of people as morally responsible agents and as objects of ethical concern. They would distract us from what some hold to be the social scientist's duty to make the world a better place through the improvement of our moral consciousness of salient facts about the social setting. This is a position on value neutrality that characterizes the critical theorists, among others (see Chapter 8, "Critical Theory").

Some proponents of the unavoidability of moral commitments in social science argue that, to begin with, there is no such thing as objectivity either in natural or in social science. Citing writers like Thomas Kuhn, they hold that the very notion has been undermined fatally by advances in epistemology or the sociology of science, or the postmodern deconstruction of science. This strategy is heir to a long tradition, according to which the distinction between facts and values is an unfounded one. It connects the issue of values in science with the agenda of the theory of knowledge: What sorts of facts can we have knowledge of? Are there facts we can know independent of interpretations, descriptions, and evaluations? Are there moral facts? Do they differ from other kinds of facts? How can we tell them apart? Either side we take on the question of value freedom in social science commits us to positions on these fundamental epistemological issues.

If we cannot propound a good account of the difference between statements of fact and expressions of moral evaluation, then the debate over whether a discipline should be value free is moot. The most popular argument against the distinction is based on the alleged impossibility of providing pure descriptions, without the implicit importation of evaluations or prescriptions.

The vocabulary of ordinary language and of the social sciences is replete with value-laden terms. For example, to describe a tribal system as primitive, a political system as a regime, an economic system as capitalist, or behavior as intelligent seems to combine description and evaluation. Even when social scientists give explicitly stipulative definitions of such terms, free from their ordinary connotations, the terms retain their "halo" of moral approval or disapproval. Thus, modern economic theory's definition of rationality as utility maximizing can be claimed to be neutral on the moral desirability of utility maximizing. But since *rational* is an ordinary term of approval, this claim carries little weight.

It is because of the halo effect of ordinary meanings that social scientists who endorse the notion of value-free social science have often had recourse to neologisms. And for their trouble, they have been accused of producing jargon that merely rephrases common sense in indecipherable circumlocutions. Their aim, of course, has been to avoid ordinary connotations. But the results in scientific advance have never seemed to justify the effort. The reason, argue proponents of value-laden social science, is that the moral dimension is an indispensable part of the explanatory strategy for rendering human affairs intelligible.

The way a social scientist selects problems to work on, the factors cited to explain behavior, and the evidence sought to substantiate these explanations all reflect the significance and meaning the social scientist attaches to them.

To focus on a particular problem is to evaluate it as more important than others, and importance is based on evaluation in the light of human values. Moreover, the terms in which events, institutions, and behavior are to be described must be meaningful to the participants in these events, institutions, and activities. But again, meaningfulness is a reflection of rules, including moral principles. A social fact cannot be identified and described in terms of the "mere" behavior of the human bodies that participate in it. It must be described from within their points of view, perhaps from the basis of the deeper meanings of these institutions.

A social science that sought to efface the moral dimension from its descriptions and explanations would simply serve the interests of some other moral conception. It would reflect values foreign to those that animate our conception of ourselves. A value-free social science, if successful in providing a predictively powerful theory of human behavior, would serve the interests of those powerful and willing enough to disregard human rights and individual autonomy. It would enable them to override meaningful action and manipulate behavior. But more likely, such a social discipline would simply be a pseudoscience (such as Nazi "racial science") serving as an empty rationalization for the socially harmful goals of the powerful. At any rate, that is what the opponent of value-free social science would argue.

One line of reply to this argument grants that social science does have some or all of these ineliminable moral dimensions. However, it goes on to identify the same or similar features in natural science. The scientist's interests also determine what phenomenon will be singled out for study, in what terms the facts will be described, how the evidence will be assessed, et cetera. This is part of what makes science a fallible enterprise: scientists are human, and what they do is as value charged as any other human activity. But that means either that value ladenness is no obstacle to scientific knowledge or that it is at least possible to reduce its obstructive effects enough to make scientific progress.

That may well be the beginning of a good argument against those who say that social science is impossible because we cannot be objective, that is, value free, in our account of our own activities. But it is no argument against the claim that social science is essentially a nonobjective enterprise, one in which progress is not measured by the standards in force among the natural sciences. Such an argument makes social inquiry of a piece with moral inquiry, so the admonition to minimize its value-laden character is a profound mistake. In this view, divesting social science of values would simply prevent us from pursuing our "science" of human action altogether.

It is clear that a full positive reply to this sort of normative argument involves little less than an entire philosophy of social science. It requires that we successfully naturalize the concepts we employ to explain human action. Short of that, the defender of value-free social science may still extol the importance of social scientists' being upfront with their evaluative commitments. This way at least others can make appropriate adjustments in their own interpretations of social claims.

But if moral commitment is a central feature of social science, then perhaps it will provide us with an explanation of why the results of social science are so different from those of natural science—and a justification for this difference as well. For few philosophers, even the most empiricist among them, have ever expected any sort of scientific progress in moral philosophy. This is indeed a discipline in which progress is never a matter of steady improvements in predictive success. Rather, moral philosophy is a matter of deepening intelligibility and coherence. If social science is really a branch of moral philosophy, perhaps the opponents of naturalism are right after all.

AMARTYA SEN ON MORAL THEORY AND SOCIAL SCIENCE

Many of the issues we have just discussed and several other ones relevant to the relation between social science and moral philosophy are effectively raised by the work of Amartya Sen, a Nobel Prize–winning theoretical and applied economist. In many works over a long period Sen has devoted himself to understanding and accelerating development in the Third World. The motivation for this work is obviously normative, so obvious that Sen doesn't even feel the need to argue for it. Everyone favors enhancing the lives of people in the Third World. This is an uncontroversial moral goal or end. It has also seemed equally obvious that the way to make people better off in developing countries is to make them richer, to raise their standards of living, to increase their average per capita income. This has made development in the Third World largely a matter for economists. Sen's importance in contemporary debates about development is in large measure the result of the fact that, though he himself is an economist, he has provided powerful arguments from economics and associated social sciences to show that development in the Third World ought not be left to economists, nor be treated as solely an economic problem.

The questions that confront students of development are twofold: first, exactly what should we aim at in development, and second, how can we most efficiently attain it? The first is plainly a question of values, a moral or normative one. The second is a factual question about the best means to attain this end. Interestingly, Sen gives the same answer to both questions. To the moral question of what we should aim at in development, Sen replies that we should aim at enhancing five distinct capacities: those fostered by political freedoms, including especially democratic party politics; economic facilities of the sort provided by free markets; guarantees of transparency in government, especially the rule of law and the absence of corruption; the protective security of a social safety net; and social opportunities free from caste, race, or gender discrimination. Capacities to live a flourishing life are enhanced by improvements along these five dimensions. Each is distinct and different, none can be derived from the others, and each individual must be free to exercise that mixture of these capabilities he or she chooses.

As to the factual question of how we can best attain this outcome, the answer Sen provides is that we can, as a matter of fact, most efficiently do so if we aim at each of the five as a means to all of the five as ends.

To show that the way to attain the normative outcome he advocates is to aim at enhancing each of the five capabilities, Sen marshals a great deal of social science—economics, political science, social psychology, anthropology. This evidence shows two distinct things: first, aiming at one of these five capacities will, under some circumstances, be part of the fastest way to attain one or more of the others, instead of aiming at the latter directly. For example, Chinese communist efforts to ensure literacy, nutrition, and health care probably enabled China much more rapidly to accelerate capitalist economic growth after the abandonment of central planning. By contrast, the much longer history of well-established economic free markets in Brazil has not eradicated widespread poverty, ill health, and other economic developmental problems. Second, there is good evidence that none of these five aims need be sacrificed to or incompatible with any other one. Thus, Sen advances evidence and argument against the widespread twentieth-century idea that economic growth requires the sacrifice of political pluralism and democratic processes. In fact, Sen has shown in a number of groundbreaking studies that democratic political institutions are the best assurance against catastrophic economic collapse that result in widespread famines. It turns out that famines are not the result of food shortages, but of the deprivation of economic entitlement guarantees that can best be ensured by democratic governments. Another of Sen's powerful empirical arguments is that enhancing women's social opportunities, through education especially, is a faster route to economic development than any policy that aims directly at facilitating and ensuring private enterprise and exchange.

Here objective, descriptive, factual social science is employed in order to identify the ways and means of development. Sen's arguments are powerful but they may be empirically disputed. However, they are not philosophically controversial.

But Sen employs important findings and theories from all the social sciences for a much more philosophically controversial purpose, one to which debate about the fact/value distinction is relevant. Sen argues for the normative goal that he advocates largely by advancing arguments against utilitarianism, Rawls's theory of justice, and other competing ethical theories. And the arguments all appeal to matters of empirical fact that it is hard to dispute. Yet, although the facts are hard to dispute, the argument that incorporates them may not be as forceful. Indeed, if there is a fact/value distinction, it will be question begging.

Sen argues that all the leading alternative moral theories—whose developmental objectives differ from his normative commitment to capability enhancement—are defective, owing to their inadequate *informational bases*. A moral theory's informational base is the set of facts and theories about human affairs that it deems relevant to choosing ends and values, rights and duties. Let's consider how this criticism works for utilitarianism and Rawls's theories.

The informational base of utilitarianism is, of course, people's welfare, their utilities, or rather the degree to which their preferences are satisfied. The only information utilitarianism requires or allows as relevant to make moral judgments is information about the welfare consequences of various attainable alternatives. It then requires those policies that maximize utility or most fully satisfy the preferences of all persons affected. Often, of course, reliable information about the welfare consequences for many people of all relevant alternatives is difficult to acquire. But Sen's objection is not based on this fact. Rather, he argues that utilitarianism ignores other morally relevant facts, including facts about preference satisfaction. In doing so, it reveals itself to be an informationally inadequate moral theory. Here is one set of findings from social psychology that utilitarianism ignores: the phenomenon of *adaptive expectations*. In brief, people's self-reported levels of welfare, happiness, and pleasure—utilitarianism's informational base—is usually a matter of adaptation to life situation. Thus, very poor Indian itinerant laborers self-report to be well satisfied under conditions most Westerners would find appalling. On the other hand, lottery winners report themselves to be no happier six months after their windfalls than they were before. When money income is substituted for self-reports as a measure of welfare or the degree to which preferences are satisfied, the results are equally disturbing: the same amount of money buys a very different quality of life depending on many different nonmonetary circumstances. Sen's bottom line for utilitarianism is that it is incapable of taking information available from social science into account in deciding on policies, even when every reasonable person will grant that the information is relevant. That is the sense in which utilitarianism's informational base is inadequate.

Sen makes the same criticism of Rawls's theory. He argues that applying the principles of justice identified by Rawls in the priority order that he requires presupposes a minimum level of well-being that in many circumstances has not yet been attained. This minimum level is identifiable in terms of measures social scientists can make of the degree to which communities have attained the capacities Sen identifies as the morally relevant ones. He offers a similar critique of other moral theories, as failing to make room in the inputs of their machinery for determining morally permissible or required outcomes of all morally relevant facts.

How can exponents of the moral theories Sen criticizes respond to his argument? One way is by invoking a strong fact/value distinction and arguing that Sen's critique begs the question against their theories. The counterargument would run as follows: the judgment that some fact uncovered by social-scientific means should be in the informational base of any theory is not itself a factual claim. Notice the operative verb in the last sentence: "should be." This makes it a normative, moral, evaluative claim. Of course the statement or description of the fact uncovered by empirical research is not itself an expression of value or a normative claim. But the insistence that a particular moral or normative theory that does not treat the fact as morally relevant *should* do so, and is morally inadequate or defective if it does not, is obviously not a factual claim at all. When Sen argues that subjective feelings of well-being are not what developmental policy should aim at, since people who are badly off have high levels of subjective welfare, he is in effect simply rejecting utilitarianism, not advancing a factual argument against it.

Of course, we may well agree with Sen about the inadequacy of the informational base of a moral theory such as utilitarianism. But this just means that we reject it, not that we have a factual or empirical argument against it. In much of his writing Sen offers a compelling case for the five capacities he identifies as being both the end or objective of development and the means to attain them. The proponents of a sharp fact/value distinction may even concur in his conclusion. But if they (and we) concur in Sen's conclusion, then they will hold not that Sen has rationally convinced us of a normative conclusion, but that he has shown us that we share his fundamental normative commitments.

How Sen approaches the fact/value distinction turns out to be crucial to how we are to understand and evaluate Sen's extremely important arguments about how development in the Third World should proceed. A powerful argument that the distinction is groundless would add greatly to the power of his objections to alternative moral theories.

FEMINIST PHILOSOPHY OF (SOCIAL) SCIENCE

Disinterested, objective science has not always been beneficial in its impact. Especially and particularly during the twentieth century, social science has persistently provided more efficient and effective ways of harming people, other organisms, and the environment. It has done so in part by providing unwarranted rationalizations for policies that effect such harms. This trend enjoins an obligation among social scientists, and others who may influence policy, to reduce as much as possible these untoward consequences. The best way to do this, some philosophers of science and of social science argue, is to make the sciences, especially the social and behavioral ones, more inclusive. This is not just a matter of increasing the numbers of women and of marginalized racial, ethnic, and sexual groups who undertake social science research. It requires, according to some, changes in the philosophy and particularly the epistemology of the social sciences—changes that reflect the value ladenness of science itself.

Since women are hardly a minority among humans, it will be especially important that their interests and values be represented in decisions about the investment of scarce resources of thought, experiment, and observation in the framing of scientific theories, especially those that are likely to affect women the most. And some feminist philosophers of science have gone on to claim that this representation requires, or at least can be enhanced by, the epistemological inclusion of women in science. These philosophers begin their examination of science from an epistemological claim, sometimes called *standpoint theory*. This theory begins with the uncontroversial thesis that certain facts are relevant to the assessment of scientific theories that are detectable only from certain points of view, or standpoints. Sometimes the point of view or standpoint in question involves using a certain apparatus; sometimes, these philosophers argue, it requires being a woman or a member of a social class or a racial minority or having a certain sexual orientation. Standpoints will be particularly important to detecting social facts, of course.

To be interesting, the thesis needs to be given strong and potentially controversial content. It needs to be understood as claiming not merely that if a male or a Caucasian or a corporate executive or a heterosexual were in the same epistemic position as women or a minority or a relevant social class, the male would detect the same fact. Rather, it must be seen as claiming that males cannot detect such a fact for the same reason they cannot be female. The fact must evidently be relatively complex, perhaps historical, certainly theoretical and not open merely to someone equipped with the five senses.

Feminist standpoint theorists have not been reluctant to identify such facts. Typically they are facts that are hard to quantify or even to fully describe in ordinary or scientific vocabularies; facts about the long-term effects of oppression, subordination, discrimination, stereotyping. These are hard facts and undeniable ones, despite all the difficulty there may be describing them, and they can be inaccessible merely from description or from a brief and/or simulated personal encounter. One has to live the standpoint to really detect the relevant facts.

Few standpoint theorists allege that physical or chemical facts are missed by failure to attend to the findings from a woman's or other marginalized standpoint, though cases have been made for the occurrence of such failures in biology. For example, it might be claimed that the initial focus of sociobiologists on evolutionarily optimal male mating strategies (maximize the number of females fertilized, minimize energy expenditure on offspring) in nonhuman species and the failure to notice female strategies (allow access to males with the best genes and a demonstrated willingness to commit resources to offspring) was owing to male biologists' incapability of locating themselves in the relevant standpoint. In all the social and behavioral sciences, especially the ones that forgo interpretation for a naturalistic approach to behavior, action, and institutions, important facts are missed through want of observation from the standpoint of women.

Opponents of standpoint theory will, of course, appeal to examples from natural science to argue against its epistemological claim. They will note that in biology all it took was for female biologists to draw the attention of their male colleagues to the facts of female mating strategies among birds for the entire discipline to revise the theory of optimal sexual strategies to accommodate the facts. This counterargument shows that what standpoint theorists need to do is very difficult: they need to identify the facts inaccessible from other standpoints in a way that forces those occupying the other standpoints to grant the existence of the facts, and to argue that these facts cannot be grasped or grasped in the same way or most accurately or most completely from these other standpoints. It remains to be seen whether this epistemological claim can be vindicated.

Standpoint theory does not exhaust the feminist philosophy of science, and in fact its sternest critics have included feminist philosophers of science who honor the aspirations of standpoint theory and seek to attain them from other premises, in particular from those congenial to naturalistic philosophies of social science. The aspirations of standpoint theory include the emancipation not just of women but of all who have suffered from the very failures of objectivity and disinterestedness that science officially may extol but that scientists actually fall short of. Feminist philosophers of social science do not need so strong an epistemological thesis as standpoint theory to identify facts that male scientists, owing to their interests, have missed. Feminist empiricists will recognize that such facts do require substantial theory to be recognized, theory that the nonscientific interests, values, even tastes of scientists brought up in a sexist world have probably prevented them from hitting upon. In the views of these feminists, the theories and the broadening of research programs to accommodate a full range of human interests may require, not just philosophical changes, but that counterevidence to theories reflecting male bias be wielded in politically effective ways.

Because feminist philosophers of science have been more attentive to developments in social science, they have emphasized the social character of research, the division of scientific labor, and the shaping of its research agenda. By contrast, the traditional philosophy of science has embraced science as the enterprise of individuals—Kepler, Galileo, Newton, Lavoisier, Darwin, Einstein. In this emphasis on individuals, it has perhaps been overly influenced by the Cartesian tradition in epistemology, which begins with Descartes's solipsistic skepticism and his consequent attempt to construct all knowledge from his own private experience. Modern science is, of course, an enterprise of teams and groups, communities and societies; indeed, of institutions and governments. Feminists have noted both the strengths and the weaknesses of this fact. On the one hand, the scientific community often serves to distribute research tasks in efficient and coherent ways, to support and to scrutinize findings and theories that individuals advance, and to provide a reward (and punishment) structure that gives scientists incentives to advance the research frontier. On the other hand, the community can be a source of prejudice, blinding individuals to empirical facts, offering perverse incentives to complicity in such ignorance, and blinding scientists to important human needs and values that should have a role in driving the direction of both pure and applied research. We need to take account of the social character of natural and social scientific inquiry and of its gendered deformation. Feminist philosophers argue that doing so should have an impact on the future of such inquiry and our philosophical assessment of it.

As we noted, empiricists usually distinguish facts from values and observe that science has long been characterized by a commitment to *value freedom*. They are ostensibly committed to not allowing the tastes, preferences, wishes, hopes, likes, dislikes, fears, prejudices, animosities, and hatreds—the values of scientists—to govern what is accepted as objective knowledge. Doing so completely and effectively, some opponents of the distinction argue, requires noncircularity in drawing the fact/value distinction. And as we have also noted, some philosophers, both feminists and nonfeminists, believe this is impossible.

But isn't the fixation of factual claims by value judgments just the sort of thing that objective, disinterested science should avoid or expunge, difficult though that may be? Of course, it does not always succeed in acting on this commitment, but science is supposed to be self-corrective: the methods of science, and in particular the control of theory by observation, are held, rightly in the eyes of feminist empiricist philosophers, to mitigate and minimize these failures. However, this is at most a negative virtue of the scientific method. At best it ensures that, in the long run, science will not go wrong epistemically. First of all, however, in the long run we are all dead. Feminist and other philosophers of science are committed along with scientists to seeing that science not go wrong in the short and the medium term, along with the long run. Second, merely avoiding error is, in their view, not enough. Avoiding error is not a motive that will explain the actual direction in which science has proceeded or how it should proceed. To explain the actual direction, at least in part, we need to identify the values of scientists, the groups and individuals, who drive it. And if we seek to change its direction, we may need to widen the range of interests represented in the scientific community.

Like all intentional human activities, scientific activity is determined not just by what we believe but also by what we want. The belief that it is raining won't send you out with an umbrella, unless you *want* to stay dry. Now, scientists don't just search for the *truth*, or even for *truths*. There is an infinite supply of the latter, and we will never make so much as a dent in the quantity of unknown truths. Science searches for *significant* truths. But what makes a statement significant and therefore worthy of scientific investigation or, for that matter, insignificant and not worthy? Feminist philosophers of science argue that the history of science is full of inquiries about statements deemed to be significant because of the values, interests, and objectives of the men who have dominated science. Likewise, many lines of inquiry are absent from its history because according to these same values, the questions they explored were insignificant. It is easy to give concrete examples of a persistent one-sidedness in according significance and insignificance to research questions. Recall the history of investigating mating strategies in evolutionary biology. Though biologists ignored female reproductive strategies in nonhumans, when it came to contraception the focus of pharmaceutical intervention was on women. On the other hand, in the treatment of depression (a disorder more frequent among women), pharmaceuticals were tested on men only, owing to the assumption that differences between male and female physiology were insignificant. Somewhere in the cognitive background of these decisions about how to proceed in science there were value judgments that neglected the interests of women.

Feminist philosophers of science have come to insist that there are in science, both natural and social, vast blind spots and blank spaces resulting from 2,500 years of male domination in identifying which questions are significant. What science needs to do now, or rather what women have always needed science to do, is to treat research questions significant to women. And the same goes for any other group, class, or race disposed in the identification of significant and insignificant research questions.

The crucial point for social science in this argument is not that judgments of significance should be forgone. Social scientists cannot do so. There are too many research questions to choose from in science's search for truths. Given scarce resources, human needs, and the importance that wonder attaches to questions, we have no alternative but to order questions by their significance *to us*. The feminist philosopher of science merely insists that we order inquiry on the basis of significance to *all of us*.

Identifying a role for value judgments in social science is not the end of the feminist agenda in the philosophy of science. In fact, it is probably closer to the beginning of it. Many feminist philosophers of social science have been interpretationalist in their views about the human disciplines. They have argued further that the real besetting sin of naturalism in social science is that of mistaking masculine styles of scientific inquiry for all scientific inquiry. Thus, they have argued, for example, that the demands for unification in scientific theorizing and explanation are often premature, counterproductive of scientific progress, or even unreasonable in a mature discipline. Feminist philosophy of science encourages pluralism. Women, and social science as they would pursue it, are more prepared than traditional male-dominated science to tolerate multiple, competing, complementary, and partial explanations without the expectation of a near-term weighting of importance, placement in a (patriarchal) hierarchy of causes, or unification under a single, complete theory. This ability to tolerate and a willingness to encourage a variety of approaches to the same problem in sociology or economics, for example, reflects women's greater sensitivity to the role of plural values—multiple judgments of significance—in driving scientific research.

Since it seems obvious that multiple assessments of significance should be encouraged by the experimental attitude of naturalistic social science, the feminist commitment to pluralism should be equally embraced by all, at the evident expense of the reductionistic proclivities of naturalism. Similarly, sensitivity to feminist discoveries about the role of values, both nefarious and benevolent, in significance decisions has implications for how the objectivity of science should be understood. Objectivity, these philosophers argue, cannot after all be a matter of complete disinterestedness, of value neutrality or detachment of the scientist from the object of inquiry. For if this were so, there would be no motivation, in judgments of significance, for the inquiry to begin with.

Some feminist philosophers of social science will make common cause with interpretationalists, rejecting the centrality of prediction and especially of control to the scientific enterprise. Their suggestion that the sciences of society and behavior should proceed in this way reflects what they hold to be masculine biases also reflected in the subordination of women and other marginalized groups. The methodology of prediction and control fails to gain the knowledge that might derive from a more cooperative relationship with the objects of scientific study, be they human or infrahuman. Among the oldest account of scientific method is Francis Bacon's seventeenth-century notion that the scientist subjects *Mother* Nature to a sort of torture in order to secure *her* secrets. Even if this is a metaphor, it may not be innocent. And there are other metaphors at work in scientific explanation that reflect a male bias harmful both to the real objectives of science and to women, independent of their purported payoff in scientific understanding.

It is not surprising that, by and large, the feminist philosophers whose work has had the most influence in the philosophy of natural science are the empiricists and naturalists. They have argued that their conclusions about how science proceeds and how it should proceed are perfectly compatible with the empiricism and naturalism that characterizes much contemporary nonfeminist philosophy of science. As noted, most feminist philosophers of social science find themselves much more in sympathy with interpretation as the goal of social science; they therefore take up an adversarial stance against naturalism and its aim of producing value-free, objective knowledge of the sort we expect from natural science. By contrast, feminist empiricist philosophers of social science do not challenge science's aim to provide objective knowledge. They seek to broaden our understanding of the role of interests and values in choosing the domains of significant inquiry. At a minimum, objectivity in social science consists in recognizing this role for values.

DANGEROUS QUESTIONS, MORAL OBLIGATIONS, AND PREDICTIVE KNOWLEDGE

Controversial subjects are the social scientist's stock in trade. A particular premium is put on social science that provides revisionist, debunking, or otherwise startling conclusions at variance with either common beliefs about the past or hopeful expectations about the future. But some social scientists and many who are not social scientists hold that some controversial questions of potential interest to social scientists ought not be pursued. For even correct answers to those questions are morally dangerous and can serve no good purpose in the guidance of social policy. Accordingly, social science should exercise a sort of self-denial, steering away from these topics.

Examples of such morally dangerous topics come readily to mind. Perhaps the most famous are a succession of studies that employed IQ tests to measure intelligence and compare average IQs between the sexes and among socioeconomic, ethnic, and racial groups. Some researchers in this area have concluded that differences in average IQ among such groups can best be explained by genetic, rather than environmental, factors. It is pretty obvious why such a conclusion might be dangerous. Regardless of what the social scientists who conduct such studies think their policy ramifications should be, others have more power over the adoption and implementation of policy. Politicians might use such findings to discourage steps to equalize the educational opportunity of all people. Even if the findings were right, such a policy would not follow from them. But they are easy to misunderstand and even easier to abuse in order to clothe racist or sexist practices in a mantle of scientific respectability. Similarly, nefarious consequences are said to follow from sociobiological speculations about the origins and character of social institutions. If sex role differences, fear of strangers, or caste and class systems are somehow written into our genetic programs, then it is widely supposed there is little we can do by altering the environment to eliminate these morally undesirable features of society. These studies thus seem a recipe for the status quo, if not for retrograde social policies.

Studies with apparently distasteful findings often provoke two sorts of reactions. The first is an examination of the scientific methods, theories, and findings that seeks to show, solely on scientific grounds, that the theories are in themselves inadequate, defective, or fundamentally confused. Philosophers have taken an especially prominent role in this enterprise and have applied the tools of the logician and the philosopher of science to the assessment of particular theories. They have scrutinized the IQ theory of general intelligence, sociobiology, and for that matter Marxian social and economic theories, which are said to have inimical effects on prospects for human freedom and economic progress.

The moral repugnance of some potential answers to questions in social science also provokes the suggestion that the questions should not be studied at all. Some inquiries, it is held, can have no morally useful function and can have only bad consequences. Constitutional guarantees of freedom of inquiry rule out no subject as illegal. Nevertheless, it is held, social scientists should deny themselves certain topics because what they uncover may be dangerous, even if it is true. Here we have an obvious parallel to the moral injunctions some have sought to impose on natural scientists. People have sought to discourage nuclear physicists from working on topics relevant to weapons production and, more recently, molecular biologists from work that may result in manipulation of human and animal genomes. Those who favor banning certain lines of research insist that scientists have a responsibility to terminate related studies if they have reason to believe that the results will be misused in the interests of injustice. There is, on this view, no blanket prohibition against certain lines of research, only a conditional one. But the conditions that would morally require such self-censorship do operate in most societies today, in their view.

This moral injunction is evidently based on a consequentialist moral theory, one that enjoins certain acts if their costs for the whole society outweigh their benefits for it. One way social scientists have opposed such injunctions against certain research is by pleading a deontologically based *right* to free inquiry. There is, of course, a tension between embracing such principles and the naturalistic methods these social scientists employ. Without debating the free-inquiry claim, let us consider how much social scientific knowledge we would need in order to justify a ban on certain kinds of research.

To know whether a certain research program is morally permissible, we need to be able to predict with some reliability the long-term consequences of its research results and their dissemination. To do that we need a substantial amount of theory about human activities and institutions. In particular we need reliable knowledge about how people respond to scientific innovations and discoveries. We also must be able to establish the initial conditions about the social contexts to which these theories are applied. And finally if we are utilitarians or consequentialists, we have to be able to calculate the net costs or benefits for society of the research program if it succeeds and if it fails.

In the absence of such knowledge, it may be argued, scientists should exercise caution. For it is better to err on the side of too much self-censorship rather than too little. If there is just a chance of some scientific finding's having a very bad net effect, then that should outweigh an equal or even a greater chance of a very good effect. But even this cautious policy still requires a vast amount of social scientific knowledge. Moreover, since we can at this point predict with accuracy almost none of the effects of scientific discoveries and their dissemination, such a cautious principle would foreclose almost every line of research—pure, applied, natural, or social. After all, almost any discovery could, for all we know, have costs that vastly outweigh its benefits.

In fact, studies aimed at acquiring the kind of social theory we would need to determine the impact of new ideas on society are themselves socially dangerous. For although they would enable us to decide whether to pursue certain issues, they would also enable those in power to manipulate social changes in directions that they might prefer in spite of their great costs to society as a whole. So perhaps the very theory we require in order to decide whether some questions should not be examined is itself such a prohibited area of inquiry.

The obverse of prohibited topics for social science is its required ones. Critical theory, for example, tells us that the aim of social science should be the emancipation of humans from bonds that restrict their freedom. The social scientist is responsible for uncovering the real meanings of social processes, institutions, events, and ideologies. Of course, that may mean violating the rights of individuals to privacy and confidentiality in their pursuit of nonemancipatory goals. Thus, whereas ordinary moral scruples will prohibit bugging a jury room, critical theory may sanction or even require it. For it might provide understanding that demythologizes this coercive social institution and thus emancipates us from the system of justice characteristic of late capitalism.

Like the prohibition against certain lines of inquiry, the prescription of some topics because of their emancipatory potential requires a great deal of social scientific knowledge. To identify topics of inquiry as potentially emancipatory requires the same knowledge of the impact of new discoveries and their dissemination on society. Otherwise, how can we tell whether uncovering hidden meanings will emancipate or whether they will be greeted with indifference? In fact, providing such a predictively successful theory about the influence of new discoveries on society as a whole is probably the first priority for an approach to social science that makes human emancipation the central goal of social science. Because of the allegedly reflexive character of social science, however, such a theory may itself be impossible. Once it comes into general circulation, its influence on human actions may lead to its own falsification. What is more serious is the notion that a philosophy like critical theory, which rejects positivism as a method in social science, may require a theory that meets positivist standards of predictive success. For only such a theory will underwrite the moral obligations that critical theory places upon social scientists.

Introduction to the Literature

N. Block and J. Dworkin, eds., *The I.Q. Controversy*, treats the interaction of methodological and normative factors that bear on whether a potentially explosive line of research should be pursued at all.

The distinction between factual and normative or evaluative descriptions goes back to David Hume, *Treatise of Human Nature*, and more recently, G. E. Moore, *Principia Ethica*. E. Nagel, *The Structure of Science*, Chapter 13, introduces the problem well and provides a strong plea for value freedom. Before him, M. Weber, *The Methodology of the Social Sciences*, argued strongly for it among social scientists. A very different view is defended in G. Myrdal, *Objectivity in Social Science*.

The anthology edited by L. I. Krimerman and the one edited by M. Brodbeck contain influential articles about the question of whether social sciences can or should be value free. A. Ryan's anthology includes an important article by Charles Taylor, "Neutrality in Political Science." This paper is also reprinted in Martin and McIntyre, *Readings in the Philosophy of Social Science*, along with excerpts from and papers by Weber and Nagel on objectivity and value-oriented biases in social inquiry. Nagel's paper is also reprinted in Steel and Guala, along with another one on related issues by Hacking. Martin and McIntyre also includes N. Weisstein, "Psychology Constructs the Female," and A. Wylie, "Reasoning About Ourselves: Feminist Methodology in the Social Sciences," which treat the implicit bias favoring males in traditional social science. An important paper by Wylie is also anthologized in Steel and Guala.

Sen's most accessible work on the relation of facts and values in human development is *Development as Freedom*.

CHAPTER FIFTEEN

Social Science and the Enduring Questions of Philosophy

This chapter reviews some of the themes of the rest of the book in order to vindicate the claim that in pursuing their disciplines social scientists must take sides on fundamental philosophical questions—matters of epistemology, metaphysics, and ethics. These are questions that their disciplines can't decide, but the answers to which make a difference for the direction and prospects of social inquiry.

SOCIAL SCIENCE AND THE ENDURING QUESTIONS OF PHILOSOPHY

The problems of the philosophy of social science are problems both for philosophy and for social science. They are problems of philosophy because their ultimate resolution turns on the response to philosophical challenges that have been with us since Plato. They are problems of social science because social scientists inevitably take sides on them, whether they realize it or not. Moreover, social scientists have defended competing and irreconcilable approaches to their own disciplines by appeal to philosophical theories. As noted in Chapter 2, even the claim that philosophical reflection is irrelevant to advancing knowledge in social science is itself a philosophical claim. Social scientists indifferent to philosophy can embrace this view. But unless they argue for it, their view must appear to others to be sheer prejudice. However, an argument for the irrelevance of philosophy is itself philosophy, whether we call it that or not.

It should not really be surprising that the social sciences and philosophy bear a profound and indissoluble link to each other. Like the natural sciences, each of the social sciences is a discipline that was once part and parcel of philosophy. Indeed, whereas the natural sciences separated themselves from philosophy in the 2,200 years from Euclid to Darwin, the social sciences became independent only during the course of the nineteenth and twentieth centuries. In separating themselves from philosophy, the natural sciences left for the continued reflection of philosophy questions that they could not deal with: What are numbers and points? What are space and time? Is there substance? It has been easy for natural scientists to leave these questions to philosophy. For they have been busy, especially in the centuries since Galileo, providing more and more detailed knowledge about large numbers of substances at widely separated points of space and time. As Thomas Kuhn noted, it has only been at periods of crisis in the development of physics or chemistry that natural scientists have turned to philosophy and taken seriously questions about the foundations of their disciplines. More often than not, the crises have been surmounted by a new piece of technology, or a new nonphilosophical breakthrough. These scientific achievements have themselves had philosophical implications.

Since Newton, advances in physical theory have had a more profound impact on our view of philosophical problems than advances in philosophy have had on the natural sciences. Natural science has forced philosophy to come to terms with materialism, mechanism, first determinism and then indeterminism, relativity, evolution by natural selection, and so forth. Each revolution in the natural sciences has generated new problems for philosophy.

But that is certainly not the case in the relationship between philosophy and social science. There has of course been much new and original in each of the social sciences. But some of these innovations have not met with the uniform acceptance of social scientists that would force philosophy to take them seriously. And the rest of these innovations have not forced philosophy to address new problems in the way natural science has. The direction of influence between philosophy and social science still seems to be from philosophy instead of toward it. We can trace the leading ideas of almost all the social sciences back to the work of philosophers in the seventeenth and eighteenth centuries. This is not just a point about intellectual history. It shows that contemporary social science is much more bound up with the philosophical tradition than is contemporary natural science.

More than ever today, social scientists seem to be interested in philosophy, especially the philosophy of science. If Kuhn is right, that is a symptom of intellectual crisis. In the heyday of behaviorism after World War II, methodological reflection was out of favor among psychologists, economists, and other social scientists inspired by their optimism. The philosophy of science was treated as the last refuge of a social scientist incapable of making a “real” contribution to the discipline. It is a matter of some irony that the confidence about the prospects for scientific progress was based on almost nothing but a philosophical theory, logical positivism, the latest version of empiricism. That doctrine goes back certainly to the Enlightenment, and probably to Plato's contemporaries.

Pessimism about a thoroughly behavioral approach to human action drew many social scientists back to a preoccupation with philosophy after about 1975. They found in the philosophy of science a number of theories ready to explain both why behaviorism failed in social science and why empiricism is inadequate anyway as a philosophy of science. But that is what another tradition in philosophy and social science had been preaching steadily at least since Hegel in the early nineteenth century.

The social scientist's preoccupation with philosophy of science seems to be another reason to identify the distinctive problem of the philosophy of social science as that surrounding the issue of progress and the allegedly invidious comparisons to natural science. But the practical concerns of the individual disciplines also make salient fundamental issues in epistemology, metaphysics, ethics, and logic.

THE UNAVOIDABILITY OF EPISTEMOLOGY

The dispute about whether the goal of social science should be predictive improvement or increasing intelligibility is fundamentally a disagreement about the nature, extent, and justification of claims to knowledge. Of course, we'd rather not have to choose between seeking improvement in prediction and making human action more intelligible. Yet insofar as what we seek in social science is knowledge, the choice is forced upon us. The demands of predictive improvement rest on a conception of knowledge as justified by its consistency with experience, and not just past experience. For it is too easy to tailor a theory to be consistent with data that are already in. A theory that can tell us about the actual world must be composed of contingent claims, which the actual world could show to be false. A body of statements that actual events could not disconfirm would be consistent with whatever happens and thus explain nothing.

If increasing our understanding of the meaning of human actions improves our predictive powers, then of course there is no conflict. The kind of knowledge that the search for meanings provides will be the same as that which predictively confirmed claims provide. But as we have seen, there are serious obstacles in the way of achieving such predictive improvements in theories that take the search for meanings seriously. We have to decide whether the obstacles are surmountable. If we decide they are not, we must face a forced choice between intelligibility and prediction. If we choose intelligibility, we are committed to a fundamentally different epistemology, one that does not require the same sort of justification for knowledge that prediction provides. Instead, the mark of knowledge that epistemology demands is some sort of certainty or necessity of connections that the mind can grasp.

Well, why not simply hold that the house of knowledge has many mansions, that there are many different sorts of knowledge? Social scientists may freely choose among them, for all are equally legitimate ways of expanding our understanding. Some social scientists are interested in knowledge that can be applied to informing social and individual policy, that can be used to predict the consequences of planning or its absence. For them, prediction is crucial, and improvements in knowledge are measured by improvements in prediction. Other social scientists have interests to which improvements in prediction are irrelevant. For them, knowledge accumulates by increasing our detailed understanding of a culture or subculture from the inside. Predictive approaches and ones aimed at interpretation are equally valid "ways" of knowing that need not compete with each other.

This view sounds like an open-minded attitude of tolerance. But it is just a way of refusing to take seriously the problems social science faces. If there really are many different forms of knowledge, all equally valid, the question must arise: what do they have in common that makes them all knowledge? After all, the term *knowledge* has to stand for something; it can't just be an arbitrary label for a heterogeneous collection of intellectual activities that have nothing in common. To suggest that religious knowledge, for instance, rests on revelation, that moral knowledge is justified by intuition, that scientific knowledge is empirical, that our knowledge of human action is based on introspective certainty, and that they are all equally legitimate shows not so much tolerance as indifference to the claims of each of these approaches. It is the attitude that anything goes, that knowledge is whatever anyone cares to assert. If a social scientist chooses to seek one of these different kinds of knowledge, there must be a reason given for this choice. Surely it cannot be merely a matter of taste whether improvable generalizations or empathetic insight into intelligibility is the aim of a social scientist's research program. It cannot be merely a matter of taste what the social scientist will count as good evidence for a theory or explanation advanced in the pursuit of inquiry. And when a social scientist chooses one goal but allows that all other epistemic goals are equally correct, she deprives her own choice of a rational foundation.

That does not mean that once we have made a choice, we should not accept or tolerate other choices and other methods as possible alternatives. For our best views of what constitutes knowledge are fallible. Having made our epistemic choice, we could be wrong. But the fallibility of our choice does not entail either that it is the wrong choice or that there is no more evidence for it than for its competitors.

If we choose to seek predictive improvement or intelligibility of our theories as the mark of knowledge, we must allow others to identify other goals, because for all we know, we might be wrong about what constitutes knowledge. But if we don't have reasons to support our choice, and perhaps also to oppose theirs, then our choice is not rationally justified.

That is what makes epistemology unavoidable for those who hold that the aim of social science is to provide knowledge. Indifference to issues of epistemology is sometimes in fact a cover for contempt. Some natural scientists, secure in their conviction about what the right methods for attaining scientific knowledge are, express great tolerance about the appropriate methods in social science. They often decline to endorse their own methods as appropriate for the study of human action and social institutions. On their view, "anything goes" in social science. But without a good reason to show that human behavior and its consequences are so different from natural phenomena that scientific methods are inappropriate for its study, this attitude is a contemptuous one. It simply disguises the view that the "soft" sciences don't provide knowledge at all, just the free play of competing speculations, which succeed each other on grounds of fashionableness instead of justification. If social science is to provide knowledge, it cannot be indifferent to what constitutes knowledge. Nor can it accept a permanent agnosticism about the claims of incompatible theories of knowledge.

SCIENCE AND METAPHYSICS

I have argued that the epistemic choice of predictive improvement as a mark of increasing knowledge must make us dissatisfied with intentional approaches to the explanation of human behavior. Similarly, an unswerving commitment to such strategies of explanation will seriously weaken the claims of prediction as an epistemic goal for social science.

Either of these alternatives raises fundamental questions about us human beings and our place in nature, questions that have always been the special province of metaphysics. For the social scientist, taking sides on these metaphysical questions seems just as unavoidable as it is for matters of epistemology. The interpretative philosophy of social science that exempts the study of humankind from the methods appropriate in the study of the rest of nature must provide an explanation of this exemption. And the naturalistic philosophy that absorbs social science into this paradigm must explain away an equally recalcitrant fact about people.

Interpretative philosophy of social science teaches that the goals of natural science are inappropriate in the study of human behavior. Another set of aims not recognized in the natural sciences must be substituted. By analyzing the way social science actually proceeds and showing that it cannot proceed in any other way, we may be able to illustrate why the goals of natural science are wrong for the study of humans. But the question is left open of why that is so. Why must the study of humans be different from every other science? It must be because of some fact about us, in particular about our minds, thoughts, consciousness, and the facts of intentionality on which interpretation trades.

If, as Descartes held, the mind is a substance quite different from the rest of nature, operating in accordance with different principles, then we have the beginnings of an explanation of why the human sciences cannot proceed in the way the study of matter does. Metaphysical differences dictate scientific differences. Descartes argued that the mind is distinct from the body on the grounds that it has properties no chunk of matter could possibly have. His most famous argument was that our minds have the property of our not being able to doubt their existence, whereas no part of our bodies, including our brains, has this feature. I can well imagine what it would be like to wake up discovering I was missing a limb or even that my skull was empty. But I cannot imagine discovering that I have no mind, for who would make this discovery if I had none? Thus my mind has a property my body lacks: indubitable existence. Accordingly, the mind cannot be part of the body.

But this dualism runs into the gravest difficulty with the evident fact that our mental states have both physical causes and physical effects. It is hard to see how something nonphysical can have such relations. For causation is preeminently a physical relation, one that involves pushes and pulls. It requires the transfer of kinetic energy, which is a function of mass and velocity—that is, matter in motion. But the interpretationalist can turn this mystery to advantage. For the impossibility of causal relations between mind and matter provides an explanation of why a predictive science of human behavior, modeled on natural science, is quite impossible: no causation—no laws; no laws—no prediction.

Of course, some will find that such an argument proves too much, for it seems to them beyond doubt that our desires and beliefs have environmental causes and behavioral effects. They may adapt Descartes's argument to a less controversial but still sufficiently strong argument against naturalism. We may grant that mental states have causes and effects, but the sort of causation involved is not physical and does not consist in generalizations we may improve in the direction of laws. Indeed, the causal relations between mind and matter are singular and irregular. But they reflect logical or conceptual relations between the intentional content of the mind, the statements describing what we believe and want, and descriptions of action. It is these conceptual connections that force a study of meanings on us as the only way to come to grips with the mind and action.

The explanatory power of such a doctrine rests in large measure on its initial metaphysical assumption that the mind is distinct from the body and not a part of the physical world. Unless interpretationalists are content to leave unexplained the distinctiveness of social scientific method, they must face the challenge of substantiating this metaphysical view.

The naturalist has the same problem in reverse. For naturalism holds that the mind is a natural object, thus explaining the appropriateness of methods drawn from the natural sciences to its study. As we saw in Chapter 4, that is no easy matter. We have as yet no plausible explanation for the most basic fact naturalism rests on: how physical matter can have intentional content, how one arrangement of matter—the brain—can represent other arrangements of physical matter. Yet if the mind is the brain, that is what our beliefs and desires will be: my belief that Paris is the capital of France must be an arrangement of neurotransmitters at the synapses of a particular part of my cerebrum. Without invoking someone or something to interpret this physical arrangement, it seems impossible to explain how it could represent some state of affairs obtaining in France, thousands of miles from my brain, involving large areas of space and complex legal facts about them. This mystery is just as great as the dualist's mystery of how nonphysical events in the mind can have physical causes and effects.

Merely announcing that the mind is the brain will not make it so. And even if the mind is the brain, we need to understand exactly how that can be, if we are to employ this bit of metaphysics in the explanation of why some methods will be more appropriate than others in the study of the mind and its effects.

It would be understandable if impatience with these matters leads some to say that how the brain represents is a matter of science, not metaphysics, and is therefore better left to scientists than philosophers. But this response simply fails to recognize that science is in fact continuous with metaphysics. Our fundamental conception of the nature of reality and our substantive study of it are on a continuum, and each heavily influences the other. Consider the impact of Newtonian mechanics on metaphysics—determinism, materialism, corpuscularism. Consider the way in which commitment to such metaphysical views led to the expansion of the domain of Newtonian science in the absence of factual evidence of determinism, materialism, corpuscularism. The explanation of the nature of reality that Newtonian metaphysics provided underwrote its scientific strategy long before the evidence for its predictive powers became overwhelming. And finally, reflect on the fact that the overthrow of Newtonian physics had equally strong ramifications for metaphysics and indeed for epistemology. The situation is the same in social science.

The role of metaphysics may be, in fact, more critical here. For if the social sciences do not have much at present to show in the way of predictive success, then we need an explanation of why they don't—and perhaps cannot—or we need an explanation of why they will ultimately provide such knowledge. Either sort of explanation so greatly transcends narrow factual matters that it must be metaphysical.

Moreover, solving the problem of how the brain actually represents requires first a solution to the puzzle of how it could possibly represent. For without a solution to the conceptual problems of intentionality, we have no hint of where to begin in searching for a solution to the factual problem of connecting psychology and neuroscience. What is more, naturalism needs to solve the metaphysical problem of representation if it is to take our intentional explanations seriously here and now, not in some happy future time when neuroscience has established itself. For in the absence of such solutions, naturalism loses out to interpretative social science as the approach most suited to the study of intentional creatures like us.

Of course, one can always opt for the view of Skinner and other materialists who just refuse to take intentional states seriously at all. Among philosophers, this view has had some currency. Though they hold no brief for the explanatory variables Skinner adopted, they agree that intentional states just have no role in adequate scientific explanations and will, in the long run, suffer the fate of notions like "phlogiston," or "demonic possession." They will simply disappear from the best explanations of behavior. Such eliminative materialists have their own metaphysical problems, distinct from those of naturalists hoping to accommodate intentional phenomena to, instead of eliminate from, the natural sciences. Perhaps the most serious of these problems is the sheer implausibility of saying our actions are not caused by our desires and beliefs, and that we don't have sensations or thoughts. This view is so implausible that its denial is often viewed as close to an a priori truth and the most basic premise of interpretative social science (see Chapters 8 and 15). In fact, eliminative materialists have tried hard to render consistent their view that such concepts will disappear from scientific explanations with our first-person convictions that we do have such intentional states. The details need not concern us here. But the argument is as much a piece of fundamental philosophy as that required to justify naturalism or interpretationalism as a method in social science.

So, all sides of the dispute about the social sciences and their goals and methods have a metaphysical mystery to deal with. Naturally, social scientists cannot be expected to cease their work and turn to the philosophy of mind. But they have taken sides on these questions by choosing methods that are underwritten by answers to these questions. They cannot pretend that the issues do not concern them and will not in the long run have an impact on the direction of research in the social disciplines.

INDIVIDUALISM AND INSTRUMENTALISM

Those who hope to skirt metaphysical issues about the mind fix the agenda of the social sciences to macrosocial facts free from psychology and individual action. They must face equally fundamental questions addressed initially in the philosophy of science and eventually in metaphysics and epistemology.

Reductionists and methodological individualists face the problem that at least some large-scale social phenomena, their descriptions and their explanations, seem resistant to explanation and description in terms of the components that make them up. This fact is hard to reconcile with the reductionism characteristic of the physical sciences. Moreover, the obvious explanation, that such phenomena somehow reflect supra-individual agencies, is difficult to accept or even make sense of if society is composed of individuals and nothing else. Therefore, individualists must search for another explanation of the resistance of social facts to reduction. One strategy is to explain away reference to irreducible wholes as a mistake. That is, however, unconvincing to those not already wedded to individualism. Another tactic is to treat macrosocial theories, not as true or false claims about the world, but as useful instruments, tools for systematizing data, and not to be taken seriously.

This approach, however, raises questions that instrumentalism has always faced in philosophy: If these instruments are so good, what is the explanation for their usefulness? And more important, why can we not produce theories that are both good as instruments and true? Are there some computational or cognitive limitations on us preventing our producing theories in social science that seem, like theories in natural sciences, to be more than just good instruments? Or are all theories natural and social merely tools for systematizing observations? Whichever move the individualist makes leads straight into the philosophy of science and thence into epistemology and metaphysics.

The holist is no better off. Holism may justify its extravagant ontology by the instrumental success of holistic theories. But it cannot rest with such justification. It too must explain how social facts, made up of the behavior of individuals, can nevertheless be distinct from individuals. Such explanations are plainly a part of metaphysics. And holism must explain how we can have knowledge of such facts when all that ever meets our eyes is the behavior of individuals. Unless holism takes such questions seriously, its position collapses into the individualist's instrumentalism and faces the same questions it does.

PHILOSOPHY AND THE MORAL SCIENCES

Probably little needs to be said to convince us that moral philosophy has a profound bearing on the social sciences and vice versa. The social sciences were, in fact, at one time known as the moral sciences, and they remain the disciplines that help us decide matters of policy, private and public. The twentieth-century trend, evinced in economics and other disciplines, of divesting the social sciences of a moral voice has never met with general agreement, and through the vicissitudes of the century, the plea for value neutrality has sometimes been reduced to the opinion of a small minority. The majority view that social science cannot be morally neutral is faced pretty directly with the matter of what moral and social prescriptions ought to be offered.

In recent years, moral philosophy has been as much a consumer or importer of theories and findings from social science as it has been a producer of and exporter to the social sciences of moral theories about what is right, good, required, prohibited, or permitted. This tendency has reflected the same doubts about a distinction between facts and values that has animated the opponents of value-free social science. There now seems little difference between the language of arguments in political philosophy and welfare economics, for instance. But the philosopher seems less constrained by economic orthodoxy. Political philosophers are prepared to consider the possibility of interpersonal comparisons and perhaps even cardinal utility, notions that have no place among modern mathematical economists. But for those theories to gain acceptance, the arguments that economics has mounted against them must be disposed of. This is certainly a task to be faced by social scientists as well as philosophers who reject the constraints of Pareto optimality.

So here the situation is reversed. Social scientists need to concern themselves with moral philosophy both because they cannot avoid ethical issues and because they may have more to say about them than we might expect. In fact, they may be able to provide the kind of information philosophy needs in order to advance and improve its own moral theories. It is, accordingly, an intellectual duty to provide this kind of help. The duty comes with the claim that social science provides, inseparably, normative and factual knowledge.

In a way, the moral responsibilities of a normatively committed social science make the classical problems of epistemology and metaphysics even more compelling. As we have seen, choosing between competing methods of pursuing social science heavily tilts our choices about moral theories. Naturalism makes a consequential theory more inviting. Antinaturalism is more sympathetic to a theory of rights and duties than to one of general welfare. So choosing between these moral points of view makes the epistemological and metaphysical problems behind the competing methods even more pressing than their purely intellectual or academic fascination might make them.

But even those who hold that social science is at its best free from value judgments and subjective impurities must face moral problems distinctive of social science. These problems are the constraints that ethics places upon our research methods, the steps we take to communicate them, their impact on others, as well as the very questions we decide to pursue as social scientists. The moral neutrality of our theories, methods, and epistemic goals, if they are indeed neutral, does not extend to us, the social scientists who pursue these goals. We make choices either self-consciously or by default. The choices seem better made as a result of serious reflection than sheer inadvertence. And such reflection takes the form of moral philosophy and applied ethics.

The first thing one learns about moral philosophy is that, like the other divisions of the subject, it too is wracked with controversy and disagreements both fundamental and derivative. Yet in contrast to the case with other areas of philosophy, we cannot remain agnostic for long about these disagreements. For they have an immediate bearing on our conduct and its effects on others and ourselves.

CONCLUSION

This introduction is meant largely for social scientists. Its aims have been three: to introduce the traditional problems of the philosophy of social science; to connect these problems with the methodological, factual, and moral choices that social scientists themselves make; and to show how the problems bring together the day-to-day research agenda of the social scientist with the most central, deepest problems of philosophy.

The first aim reflects my belief that current controversies in the philosophy of social science are almost always new versions of traditional debates. Sometimes it is difficult to recognize this fact because the jargon has changed and the participants themselves too often mistakenly think they have discovered a new issue. Today's argument between interpretational social science and naturalistic social science reflects the same issues that were debated among Weber and Durkheim, Dilthey and Comte, Mill and Marx, Hegel and Hobbes. That does not mean that current disputes are condemned to perpetual gridlock. Rather, it means that traditional insights bear a continuing relevance.

The second aim of this book is to demonstrate that relevance by showing that social scientists take sides in these disputes, whether or not they realize it. And sometimes they inadvertently take both sides, an intolerable result when the sides are mutually incompatible. For example, a naturalist cannot offer functional explanations while holding that there is no underlying causal mechanism to underwrite the explanation. An interpretationalist cannot advocate a particular policy that reflects our empathetic understanding of action while denying that prediction of the policy's effects is relevant to assessing it.

The third aim, of making the social scientist see the seriousness and the relevance of questions that daunted Plato, Descartes, Hume, Kant, and their discipline, reflects the conviction that the search for knowledge is all of one piece. But this conviction is also the basis of another aim, one that could animate an introduction to the philosophy of social science. This is the aim of encouraging philosophers to recognize the bearing of work in the social sciences to their traditional concerns. For if social scientists take sides on philosophical issues in their work, then the findings, theories, and methods of these disciplines must test, and eventually inform, the thinking of philosophers.

Bibliography

The following list contains all the books mentioned in the Introduction to the Literature sections that follow the chapters, along with some others worth studying by interested students. It is not a complete bibliography of the subject. Guidance to further work in the literature can be found in the bibliographies of the principal anthologies of the field, especially Daniel Steel and Francesco Guala's *The Philosophy of Social Science Reader*, and *Readings in the Philosophy of Social Science*, edited by M. Martin and L. C. McIntyre. Additionally, there are several journals that publish papers in the philosophy of social science regularly. Among them the most prominent are *Philosophy of Science*, *Philosophy of Social Science*, *British Journal for the Philosophy of Science*, *Economics and Philosophy*, and *Biology and Philosophy*.

Akerlof, G. A. "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* (August 1970).

Alexander, R. *Darwinism and Human Affairs*. Seattle: Washington University Press, 1979.

Axelrod, R. *The Evolution of Cooperation*. New York: Basic Books, 1984.

Beauchamp, T., Faden, R., Wallace, J., and Walters, L., eds. *Ethical Issues in Social Science Research*. Baltimore: Johns Hopkins University Press, 1982.

Becker, G. *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press, 1976.

Berger, P., and Luchman, T. *The Social Construction of Reality*. New York: Doubleday, 1966.

Bicchieri, C. *The Grammar of Society*. Cambridge: Cambridge University Press, 2006.

Blaug, M. *Economic Theory in Retrospect*. Cambridge: Cambridge University Press, 1978.

———. *The Methodology of Economics*. Cambridge: Cambridge University Press, 1980.

Block, N., ed. *Readings in the Philosophy of Psychology*, vol. 1. Cambridge, MA: Harvard University Press, 1980.

Block, N., and Dworkin, J., eds. *The I.Q. Controversy*. New York: Random House, 1976.

Bohman, J. *New Philosophy of Social Science*. Cambridge, MA: MIT Press, 1991.

Boyd, R., and Richerson, P. *Not by Genes Alone*. Chicago: University of Chicago Press, 2004.

Braybrooke, D. *Philosophy of Social Science*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

Braybrooke, D., ed. *Philosophical Problems of the Social Sciences*. New York: Macmillan, 1965.

Brodbeck, M., ed. *Readings in the Philosophy of Social Sciences*. New York: Macmillan, 1968.

Brown, R. *The Nature of Social Laws*. Cambridge: Cambridge University Press, 1984.

Buchanan, A. *Ethics, Efficiency and the Market*. Totowa, NJ: Rowman and Allanheld, 1985.

Burgess, R., and Bushell, D. *Behavioral Sociology*. New York: Columbia, 1969.

Campbell, D. *Quasi-Experiments*. San Francisco: Jossey-Bass, 1981.

Charlesworth, J. C., ed. *The Limits of Behavioralism in Political Science*. Philadelphia: American Academy of Political and Social Science, 1962.

Chomsky, N. "Review of B. F. Skinner, *Verbal Behavior*." *Language* 35 (1959): 26–58. Reprinted in N. Block, *Readings in the Philosophy of Psychology*.

Christensen, S., and Turner, D. *Folk Psychology and the Philosophy of Mind*. Hillsdale, NJ: Erlbaum, 1993.

Churchland, P. "Eliminative Materialism and the Psychological Attitudes." *Journal of Philosophy* 78 (1981): 67–90.

———. "The Logical Character of Action Explanations." *Philosophical Review* 79 (1970): 214–236.

———. *Matter and Consciousness*. Cambridge, MA: MIT Press, 1984.

———. *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press, 1979.

Cohen, G. *Karl Marx's Theory of History: A Defence.* 2nd ed. Oxford: Oxford University Press, 2001.

Collingwood, R. G. *The Idea of History*. Oxford: Oxford University Press, 1946.

Cosmedes, L., Tooby, J., and Barkow, J., eds. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press, 1992.

Cowie, F. *What's Within? Nativism Reconsidered*. New York: Oxford University Press, 1999.

Cummins, Robert. "Functional Analysis." *Journal of Philosophy* 72 (1975): 741–764.

Davidson, D. *Essays on Actions and Events*. Oxford: Oxford University Press, 1980.

Dennett, D. C. *Brainstorms*. Cambridge, MA: MIT Press, 1978.

———. *Content and Consciousness*. London: Routledge and Kegan Paul, 1969.

———. *Darwin's Dangerous Idea*. New York: Simon & Schuster, 1995.

Dray, W. *Law and Explanation in History*. Oxford: Oxford University Press, 1957.

———. *The Philosophy of History*. Englewood Cliffs, NJ: Prentice-Hall, 1964.

Dretske, F. *Explaining Behavior*. Cambridge, MA: MIT Press, 1988.

Durkheim, E. *The Rules of the Sociological Method*. New York: Free Press, 1965.

———. *Suicide*. New York: Free Press, 1951.

Easton, D. "The Meaning of Behavioralism in Political Science." In J. C. Charlesworth, *The Limits of Behavioralism in Political Science*. Philadelphia: American Academy of Political and Social Science, 1962.

Elster, J. *Making Sense of Marx*. Cambridge: Cambridge University Press, 1985.

Elster, J. *Nuts and Bolts for Social Science*. Cambridge: Cambridge University Press, 1989.

Elster, J., ed. *Rational Choice*. New York: New York University Press, 1986.

Erwin, E. *Behavior Therapy: Scientific, Philosophical and Moral Foundations*. Cambridge: Cambridge University Press, 1978.

Fiske, D. W., and Shweder, R. A. *Metatheory in Social Science*. Chicago: University of Chicago Press, 1986.

Fodor, J. *The Language of Thought*. Cambridge, MA: Harvard University Press, 1979.

———. *Representations*. Cambridge, MA: MIT Press, 1981.

———. *A Theory of Content*. Cambridge, MA: MIT Press, 1990.

Frankena, W. K. *Ethics*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1973.

Freud, S. *New Introductory Lectures on Psychoanalysis*. New York: Norton, 1933.

Friedman, M. "The Methodology of Positive Economics." In *Essays in Positive Economics*. Chicago: University of Chicago Press, 1953. Reprinted in D. Hausman, ed., *The Philosophy of Economics*, and A. Ryan, ed., *The Philosophy of Social Explanation*.

Garfinkel, H. *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall, 1967.

Geertz, C. *The Interpretation of Cultures*. New York: Basic Books, 1973.

Geuss, R. *The Idea of Critical Theory*. Cambridge: Cambridge University Press, 1981.

Giddens, A. *Sociology: A Brief but Critical Introduction*. New York: Harcourt, Brace, & Jovanovich, 1982.

Godfrey-Smith, P. *Theory and Reality*. Chicago: University of Chicago Press, 1999.

Grenfell, M. *Pierre Bourdieu: Key Concepts*. London: Acumen Press, 2008.

Grunbaum, A. *The Foundations of Psychoanalysis*. Berkeley: University of California Press, 1985.

Gutting, G. *The Cambridge Companion to Foucault*, 2nd ed. Cambridge: Cambridge University Press, 2005.

———. *Michel Foucault's Archaeology of Scientific Reason*. Cambridge: Cambridge University Press, 1989.

Gutting, G., ed. *Paradigms and Revolutions*. South Bend, IN: University of Notre Dame Press, 1980.

Habermas, J. *Knowledge and Human Interests*. Boston: Beacon Press, 1971.

Hampshire, S. *Thought and Action*. New York: Viking Press, 1959.

Harding, S. *The Science Question in Feminism*. Ithaca: Cornell University Press, 1986.

Harman, G. *The Nature of Morality*. Oxford: Oxford University Press, 1977.

Harré, R., and Secord, P. *The Explanation of Social Behavior*. Oxford: Blackwell, 1972.

Harris, M. *Cultural Materialism*. New York: Random House, 1979.

Hausman, D. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press, 1991.

Hausman, D., ed. *The Philosophy of Economics*. Cambridge: Cambridge University Press, 1984.

Hempel, C. *Aspects of Scientific Explanation*. New York: Free Press, 1965.

———. *Philosophy of Natural Science*. Englewood Cliffs, NJ: Prentice-Hall, 1966.

———. "Rational Action." Proceedings of the American Philosophical Association, 1962.

Henderson, D. *Interpretation and Explanation in the Human Sciences*. Albany: SUNY Press, 1993.

Hobbes, T. *Leviathan*. Indianapolis: Bobbs-Merrill, 1958.

Hollis, M. *The Philosophy of Social Science*. Cambridge: Cambridge University Press, 1994.

Homans, G. *The Human Group*. Cambridge, MA: Harvard University Press, 1951.

Homans, G., and Schneider, D. *Marriage, Authority and Final Causes*. New York: Free Press, 1955.

Houlgate, S. *An Introduction to Hegel: Freedom, Truth and History*, 2nd ed. Oxford: Blackwell, 2005.

Hume, D. *Enquiry Concerning Human Understanding*. Oxford: Oxford University Press, 1975.

———. *Treatise of Human Nature*. Oxford: Oxford University Press, 1888.

Humphrey, L. *Tearoom Trade*. Chicago: Aldine, 1975.

Joyce, R., *The Evolution of Morality*. Cambridge, MA: MIT Press, 2006.

Kant, I. *Foundations of the Metaphysics of Morals*. Indianapolis: Bobbs-Merrill, 1959.

Kitcher, Patricia. *Freud's Dream: A Complete Interdisciplinary Science of Mind,* Cambridge, MA: MIT Press, 2004.

Kitcher, Phillip. *Vaulting Ambition*. Cambridge, MA: MIT Press, 1986.

Knorr-Certina, K. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon Press, 1981.

Koopmans, T. *Three Essays on Economic Science*. New York: McGraw Hill, 1957.

Krimerman, L. I., ed. *The Nature and Scope of Social Science*. New York: Appleton-Century-Crofts, 1969.

Kripke, S. *On Rules and Private Language*. Cambridge, MA: Harvard University Press, 1984.

Kuhn, T. S. *Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.

Laland, K., and Brown, G. *Sense and Nonsense: Evolutionary Perspectives on Human Behavior*. New York: Oxford University Press, 2002.

Lane, J. F. *Pierre Bourdieu: A Critical Introduction*. London: Pluto Press, 2000.

Latour, B., and Woolgar, S. *Laboratory Life: The Social Construction of Scientific Facts*. Beverly Hills, CA: Sage, 1979.

Lévi-Strauss, C. *Structural Anthropology*. New York: Harper, 1968.

Little, D. *Varieties of Social Explanation*. Boulder, CO: Westview Press, 1991.

Luce, R. D., and Raiffa, H. *Games and Decisions*. New York: Wiley, 1957.

Lukes, S. *Marxism and Morality*. Oxford: Oxford University Press, 1987.

Mackie, J. L. *Ethics: Inventing Right and Wrong*. London: Penguin Books, 1979.

Malinowski, B. *A Scientific Theory of Culture*. Chapel Hill, NC: University of North Carolina Press, 1944.

Martin, M., and McIntyre, L. C., eds. *Readings in the Philosophy of Social Science*. Cambridge, MA: MIT Press, 1994.

Maynard-Smith, J. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press, 1982.

McCarthy, T. *The Critical Theory of Jurgen Habermas*. Cambridge, MA: MIT Press, 1978.

———. *Ideals and Illusions*. Cambridge, MA: MIT Press, 1991.

Melden, A. *Free Action*. London: Routledge and Kegan Paul, 1961.

Merton, R. K. *Social Theory and Social Structure*. New York: Free Press, 1957.

Milgram, S. *Obedience to Authority*. New York: Harper, 1974.

Mill, J. S. *A System of Logic*. London: Macmillan, 1866, and subsequent editions.

———. *Utilitarianism*. Indianapolis: Hackett, 1974.

Moore, G. E. *Principia Ethica*. Cambridge: Cambridge University Press, 1907.

Myrdal, G. *Objectivity in Social Science*. London: Duckworth, 1970.

Nagel, E. *The Structure of Science*. Indianapolis: Hackett, 1979.

Nagel, T. *The View from Nowhere*. Oxford: Oxford University Press, 1986.

Natanson, M., ed. *Philosophy of Social Science: A Reader*. New York: Random House, 1953.

Needham, R. *Structure and Sentiment*. Chicago: University of Chicago Press, 1959.

Nozick, R. *Anarchy, State and Utopia*. New York: Basic Books, 1974.

Olson, M. *The Logic of Collective Action*. Cambridge, MA: Harvard University Press, 1965.

Papineau, D. *For Science in the Social Sciences*. New York: St. Martin's, 1978.

Parsons, T. *The Social System*. Glencoe, IL: Free Press, 1951.

Peters, R. S. *Concept of Motivation*. London: Routledge and Kegan Paul, 1958.

Pinkard, T. *Hegel: A Biography*. Cambridge: Cambridge University Press, 2000.

———. *Hegel's Phenomenology: The Sociality of Reason*. Cambridge: Cambridge University Press, 1994.

Pippin, R. *Hegel's Idealism: The Satisfactions of Self-Consciousness*. Cambridge: Cambridge University Press, 1989.

———. *Hegel's Practical Philosophy: Rational Agency as Ethical Life.* Cambridge: Cambridge University Press, 2008.

———. *Idealism as Modernism: Hegelian Variations*. Cambridge: Cambridge University Press, 1997.

Plato. *Phaedo*. Translated by D. Gallop. Oxford: Oxford University Press, 1975.

Popper, K. *The Open Society and Its Enemies*, vols. 1 and 2. London: Routledge and Kegan Paul, 1962.

———. *The Poverty of Historicism*. London: Routledge and Kegan Paul, 1957.

Putnam, H. *Meaning and the Moral Sciences*. London: Methuen, 1978.

Quine, W. V. O. *Word and Object*. Cambridge, MA: MIT Press, 1960.

Rabin, Matthew. "Psychology and Economics." *Journal of Economic Literature* 36 (March 1998): 11–46.

Rachlin, H. *Introduction to Modern Behaviorism*. San Francisco: Freeman, 1970.

———. "Maximization Theory in Behavioral Psychology." *Behavioral and Brain Sciences* 4 (1981): 371–418.

Radcliffe-Brown, A. R. *Methodology in Social Anthropology*. Chicago: University of Chicago Press, 1958.

Rawls, J. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.

Riker, W., and Ordeshook, P. *Introduction to Positive Political Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1973.

Rorty, Richard. *Contingency, Irony, and Solidarity*. Cambridge: Cambridge University Press, 1989.

Rosenberg, A. *Economics—Mathematical Politics or Science of Diminishing Returns?* Chicago: University of Chicago Press, 1993.

———. *Microeconomic Laws: A Philosophical Analysis*. Pittsburgh: University of Pittsburgh Press, 1976.

———. *The Philosophy of Science: A Contemporary Introduction*, 3rd ed. London: Routledge, 2011.

———. *Sociobiology and the Preemption of Social Science*. Baltimore: Johns Hopkins University Press, 1979.

———. *Structure of Biological Science*. Cambridge: Cambridge University Press, 1985.

Rudner, R. *Philosophy of Social Science*. Englewood Cliffs, NJ: Prentice-Hall, 1966.

Ryan, A., ed. *The Philosophy of Social Explanation*. Oxford: Oxford University Press, 1973.

Ryan, A. *Philosophy of Social Sciences*. London: Macmillan, 1970.

Sahlins, M. *Culture and Practical Reason*. Chicago: University of Chicago Press, 1976.

Schelling, T. *Micromotives and Macrobehavior*. New York: Norton, 1978.

Schmitt, R. *Introduction to Marx and Engels*. Boulder, CO: Westview Press, 1987.

Schweder, R. A., and Levine, R. A., eds. *Culture Theory: Essays on Mind, Self and Emotions*. Cambridge: Cambridge University Press, 1984.

Searle, J. *Intentionality*. Cambridge: Cambridge University Press, 1983.

———. *Making the Social World.* Oxford: Oxford University Press, 2010.

———. "The Word Turned Upside Down." *New York Review of Books* 30 (1983): 79–83. Available online at www.nybooks.com/articles/article-preview?article_id=6083.

Sen, A. *Choice, Welfare and Measurement*. Cambridge, MA: MIT Press, 1982.

———. *Development as Freedom*. New York: Anchor, 2000.

Singer, Peter, *Marx: A Very Short Introduction*, Oxford: Oxford University Press, 2000.

Skinner, B. F. *Beyond Freedom and Dignity*. New York: Bantam, 1971.

———. *Science and Human Behavior*. New York: Macmillan, 1953.

Skyrms, B. *Choice and Chance*. Belmont, CA: Dickenson, 1966.

———. *The Evolution of the Social Contract*. Cambridge: Cambridge University Press, 1999.

———. *The Stag Hunt and the Evolution of Social Structure*. New York: Cambridge University Press, 2003.

Smelser, N., and Warner, S. *Sociological Theory*. Morristown, NJ: General Learning Press, 1976.

Smith, A. *The Wealth of Nations*. London: Penguin, 1970.

Sober, E., ed. *Conceptual Issues in Evolutionary Biology*. Cambridge, MA: MIT Press, 1983; 2nd ed., 1993, 3rd ed., 2008.

Sober, E. *The Nature of Selection*. Cambridge, MA: MIT Press, 1984.

———. *Philosophy of Biology*. Boulder, CO: Westview Press, 1993.

Sober, E., and Wilson, D. S. *Unto Others*. Cambridge, MA: Harvard University Press, 1999.

Sosa, E., and Tooley, M. *Causation*. Oxford: Oxford University Press, 1993.

Steel, D., and Guala, F. *The Philosophy of Social Science Reader*. London: Routledge, 2011.

Sterelny, K. *Thought in a Hostile World*. Oxford: Blackwell, 2003.

Stich, S. *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press, 1983.

Stich, S., and Warfield, T., eds. *Mental Representation*. Oxford: Blackwell, 1994.

Taylor, C. *Explanation of Behavior*. London: Routledge and Kegan Paul, 1964.

Taylor, R. *Action and Purpose*. Englewood Cliffs, NJ: Prentice-Hall, 1966.

Thompson, P. *Issues in Evolutionary Ethics*. Albany: SUNY Press, 1994.

Turner, J. *Theoretical Principles of Sociology, V. One: Macrodynamics*. New York: Springer, 2010.

Turner, J., and Stets, J. *The Sociology of Emotions*. Cambridge: Cambridge University Press, 2005.

Weber, M. *The Methodology of the Social Sciences*. Glencoe, IL: Free Press, 1949.

Wilson, E. O. *On Human Nature*. Cambridge, MA: Harvard University Press, 1978.

———. *Sociobiology*. Cambridge, MA: Harvard University Press, 1975.

Winch, P. *The Idea of a Social Science*. London: Routledge and Kegan Paul, 1958.

Wittgenstein, L. *Philosophical Investigations*. Oxford: Blackwell, 1953.

Wolff, R. P. *Understanding Marx*. Princeton: Princeton University Press, 1984.

Wood, A. *Hegel's Ethical Thought*. Cambridge: Cambridge University Press, 2000.

Wright, L. *Teleological Explanations*. Berkeley: University of California Press, 1976.

Index

A priori, 30
  truth, 166, 263, 300
  knowledge, 265
Action
  biological approach to, 246, 249–250
  causes of, 59, 79, 127
  deep meaning of, 160, 202
  economic, 91
  explanation of, 40, 50, 73–74, 81–82, 87, 100–103, 122, 201
    as meaningful, 119–120, 123, 125
    in terms of beliefs and desires, 37–39, 42–43, 45, 48, 72, 75, 77–78, 89, 119, 173
    in terms of causes, 42, 44–46, 49, 55–57, 59–60, 128–129
  descriptions of, 59–60, 67–68
  distinguished from behavior, 35–36, 40, 64–65, 80
  Hegel's approach to, 140–141
  intentional, 64–66
  intensionality of, 65, 67
  meaning of, 30–32, 53, 135, 160–162, 166, 173–174, 201, 295
  prediction of, 31, 51–52, 83, 101
  scientific theory of, 117–118, 278
  skepticism about, 164
  understanding of, 71, 304
  voluntary, 39, 133
Adaptation, 77, 203–204, 208–210, 231, 237, 239–241, 249–250
  cultural, 245
  of human affairs, 195–197
  of institutions, 201, 244
Aggregation, 25, 36, 170, 181, 221, 247, 262
Alienation, 148–149, 203
Altruism, 96, 199, 210–211, 220, 230
Âme collective. *See* group mind.
Animals, 76, 81, 83, 85, 87, 203, 209, 215, 226, 254
Anomie, 199, 203
Anthropologists, 1–2, 27, 30, 65, 125, 130, 165, 181
Anthropology, 94, 153
  biological, 220, 237
  cross-cultural, 135
  philosophical, 139, 157, 159, 169, 180, 192
Antinaturalism, 31, 303
Antipositivists, 152
Applicability, 97, 239
Applied ethics, 303
Ariew, A., 236
Aristotle, 22
  science, 75, 80
  theory, 23
Arms-race, 249–250
Artificial intelligence, 185
Assumptions, 5, 23, 37, 48–49, 66, 75, 93–94, 114, 248
  fundamental, 16,
  rational choice, 98, 100–102, 104–107, 112, 267
  unrealistic, 53, 109–110, 267
Astronomy, 151
Atheists, 75
Auden, Wystan Hugh, 253
Autonomy, 155, 172, 176, 180–183, 185–189, 194, 256–260, 262–264, 278
Axelrod, Robert, 216–218
Axioms, 15–16

Bacon, Francis, 288
Behavior
  and its causes and effects, 19, 262–263, 297

Behavior (*continued*)
- biological, 76, 246, 250
- causal explanation of, 44, 60, 126, 266
- causal theory of, 29–31
- choice, 102
- consistent, 100
- cooperative, 226
- deep meanings of, 27, 144–145
- distinguished from action, 35–36, 40, 64–65, 80
- explanation of, 12, 20–21, 25–26, 122–124, 173
- individual, 112, 143, 153, 178–180, 185, 189, 197, 201, 243
  - aggregating, 170–171, 175
  - and social facts, 175, 186, 194, 200, 302
- irrational, 53, 66
- homosexual, 255
- learned, 85, 209, 222
- neurotic, 144, 146
- nonverbal, 49
- operant, 80–84, 86–87
- prediction of, 45–46, 51–52, 68, 104–106, 125, 278, 298
- social, 123, 133, 156

Behavioral psychologists, 44
Behavioral science, 2, 3, 5, 8, 50, 61, 67, 189, 197, 221, 237, 239, 284
Behavioralism, 73
Behavioralist. *See* Behavoralism.
Behaviorism, 22, 71, 74, 87–88, 98, 101, 104–105, 161, 294–295
- philosophical, 74, 111

Behaviorists, 20, 22, 60, 71–73, 77–80, 83–84, 89, 227
Beliefs
- about causes, 46
- and desires, 62–67, 80, 119, 131–133, 142–144, 153, 194, 264, 299
  - as causal, 77, 118–120, 126, 300
- as an explanation for action, 37–39, 42–43, 45, 48, 72, 75, 77–78, 89, 119, 166, 173
- as connected to desires and actions, 39–42, 50, 56–57, 59, 87, 103, 120, 123, 127, 134
- categories of, 21–22
- false, 124, 144, 160, 175–176, 263
- probability, 103
- true, 90, 262
- well-justified, 114

Billiard ball model, 109–110
Biological
- anthropologist, 220, 237
- behavior, 76, 250
- functions, 77, 231
- models, 30
- processes, 32
- science, 2, 4–5, 67, 151, 184, 254
- theory, 180, 185–186, 208, 245

Biologist, 21, 77, 184, 186, 209, 211–212, 235, 249
- evolutionary, 231, 251
- molecular, 290
- population, 236

Biology, 18, 76–77, 151, 183, 185, 187, 189, 195, 200, 203, 212, 249
- evolutionary, 83, 235, 251, 286
- folk, 226
- molecular, 32, 67, 186, 243, 249
- philosopher of, 237

Blind variation, 21, 76–77, 195, 203–204, 208, 239
Blinks, 35–36
Body, 25, 35, 49, 56, 127–129, 132
- relation to mind, 55, 67–68, 71, 111, 298–299

Bourdieu, Pierre, 153, 155–157, 180
Bourgeois
- democracy, 147
- economic theory, 161, 163
- microeconomic theory, 150

Boyd, Richard, 244–245
Brain, 61–62, 79, 85–87, 127, 146, 222–223, 225, 234, 242–243, 299–300
- developmental plasticity of, 237–238
- state, 61, 67–68, 172, 183–184, 189

British empiricist philosophers, 16, 222, 260

Calvinists, 37
Capitalism, 37, 148–150, 291
Capitalist, 200, 202, 269, 277, 280
- society, 30, 149–150

Capitalistic exploitation, 150
Cardinal, 257
- numbers, 91
- utility. *See* Utility
- utility theory, 91, 97–98, 162

Cardinally measurable, 95–98, 261

Categorical imperative, 263, 265
Catholic Church, 37, 164, 196
Catholics, 177–178
Causal, 78–79, 113, 120, 129–130, 134, 264, 266
  chain, 146, 182
  claim, 13, 55–59, 118, 157, 246–247
  inquiry, 28, 55, 59, 125
  knowledge, 13–14, 30, 46, 126, 163
  law, 28, 42–44, 49–52, 56, 66, 117, 121–122, 127–128, 265
  mechanism, 15, 24–25, 76–77, 123, 127, 142, 250, 304
  regularity, 19, 68, 119, 128, 247
  relation, 45, 248, 298–299
  role, 182, 204
  theory, 27, 31, 77, 82, 118, 149, 152, 203
Causation, 13–14, 38, 43, 46, 64, 117, 128, 151, 246, 298
  metaphysical problems of, 30
  teleological, 152
Ceteris paribus clause, 40, 42, 51, 90, 102, 108, 246–248
Chemical law, 15
Chemistry, 16, 23, 67, 145, 172, 179–180, 183, 185–186, 251, 274, 294
Chomsky, N., 225, 238
Clark Kent, 62–63
Cogito, 166
Cognitive, 19, 87, 187, 205, 221, 226, 230, 238, 245, 250, 267, 286, 301
  capacities, 234, 239
  models, 225
  science, 111–112, 115
  states, 61
Collective consciousness, 171, 202. *See also* group mind
Common law, 197, 259
Common sense, 21, 26, 36, 38, 64, 89, 128, 148, 152, 197, 277
Commonsense, 20
  theory, 36, 165
  explanation, 37, 50, 79
Compliance conditions, 173
Computational, 223, 301
Comte, Auguste, 304
Conceptual schemes, 24, 165–166, 168
Conditioning, 22, 80–82
Condorcet, Marquis de, 20
Confidentiality, 256, 258, 291
Conflict-resolving institutions, 198
Consciousness, 56, 142, 148–149, 171, 255, 276, 298
Consequentialism, 261–263, 270
Constant conjunction, 13–14
Constructivist, 131, 134–135
Cooperation, 205, 210–212, 215–216, 218, 220–221, 223–227, 237–238
Cooperative
  dispositions, 218
  groups, 212
  institutions, 210–211
  strategy, 220, 223, 226, 245
Covering-law theory, 15, 46
Critical theory, 157–163, 291
Cultural, 153, 156–157, 166, 197, 209
  adaptation, 245
  anthropologists, 27, 30, 65, 135, 165
  evolution, 210, 237–239, 241, 243–245
  factors, 139
  norms, 210
  relativism, 137, 165
Culture, 2, 26–27, 30, 131–132, 135, 141, 144, 156–157, 165–166, 170, 205, 210, 212, 217–218, 224, 296
  Darwinian theory of, 238–244
  theories about, 139, 162
Cut the cake, 218, 220

Darwin, Charles, 31–32, 82, 208–209, 211, 226, 242, 285, 294
Darwinian, 201, 230–231
  biologist, 213
  cultural evolution, 239, 245
  dynamics, 209, 220
  mechanism, 250
  process, 203
  theory, 5, 76–77, 172, 204–205, 210, 212, 227, 233–234, 237, 239–241, 243–244, 249 (*see also* Evolution: theory of)
Darwinism, 236, 244–245
Dawkins, Richard, 241–242
Declaration of Independence, 263
Deconstruction, 31, 166, 277
Deductive nomological, 15, 46
Deep meanings, 137, 140–142, 144–146, 148, 152, 160, 201
Definitions, 27, 38, 53, 55–59, 74, 236, 277
Demography, 25
Dennett, Daniel, 69, 242

Deontology, 263, 270
Descartes, René, 166, 285, 298, 304
Descent of Man (Darwin), 209–210, 242
Description, 39, 107, 127–128, 151, 160, 171–175, 184, 186, 197, 273, 299, 301
  and prescription, 274, 277, 282
  intentional, 62–65, 67–68, 86
  of behavior, 74, 84, 101, 173, 262
  of belief, 57–59, 62, 77
  of initial conditions, 15
  teleological, 76
  value-laden, 275
Design problem, 223, 249–251
Desires, 43, 45, 49, 75, 79, 95, 102, 103
  basic, 164
  category of, 21
  conscious, 142
  intentional, 60, 62, 64
  ordering, 90
  unconscious, 38, 143, 144
  See also Beliefs: and desires
Determinism, 150, 155, 264, 294, 299–300
  genetics, 233-236
Dilthey, Wilhelm, 304
Dispositions, 156, 209–210, 218, 221, 230
Doctrine of "organic wholes," 179
Dr. Pangloss, 195
Dualism, 188, 298
Durkheim, Émile, 5, 20, 22, 153, 155, 170, 172, 179–182, 187, 191, 198–199, 201–203, 208, 221, 304
  Suicide, 12, 176–178, 202
  See also Group mind
Duties, 2, 133, 147, 198, 260, 263–265, 268, 281, 303. *See also* Rights
Dynamical laws of mechanics, 151

Econometrics, 25
Economic
  actions, 91
  agents, 42, 90, 94, 96, 98, 102, 104, 261, 267
  analysis, 97
  behavioral, 113
  choices, 111
  class, 141, 149, 156
  development, 280
  equilibrium, 92–95, 105, 107–108, 114, 223
  factors, 139
  models, 30, 108, 267
  organization, 37
  theory, 73, 92, 98–101, 103–104, 106, 109, 114, 213, 257, 269, 276–277
    bourgeois, 160–161, 163
    Marxian, 146, 153, 289
Economists, 30, 36, 42, 89, 91–92, 106, 110, 113, 212, 220, 260, 267
  contemporary, 5
  marginalist, 93
  mathematical, 94, 98, 302
Edgeworth, Francis Ysidro, 162
Egalitarian, 162, 268
Ego, 143–145, 199, 205, 209, 213
Einstein, Albert, 16, 23–24, 285
Electromagnetism, 183
Electron, 15, 17, 128, 171, 174–175, 187
Eliminative materialists, 300
Emotion, 226–227, 230, 275
Empirical, 139, 222
  enquiry, 158
  evidence, 56, 65
  fact, 178, 281, 285
  generalizations, 12, 19, 25–26
  investigations, 134
  knowledge, 53, 265, 296
  meanings, 16
  methods, 13
  phenomena, 15
  research, 56, 282
  science, 4, 31, 157, 165, 254
  social scientist, 31, 69, 159, 251, 266
  studies, 234, 257
  testing, 68
Empiricism, 17, 24, 30, 164, 288, 295
  logical, 14. *See also* Logical Positivism.
Empiricist, 151, 155, 176, 183, 246, 249, 288
  approach to psychological theory, 68
  British, 222
  epistemology, 17, 24, 163
  feminist, 285–286
  philosophy, 13, 162
  philosophy of science, 17, 33, 157
  scientists, 79
  social scientists, 247
  theory of knowledge, 16
End, 74, 85, 91, 264, 281. *See also* Teleology
Enlightenment, 23, 55, 154, 160, 295
Environmental, 224–225
  causes, 298
  conditions, 72, 222, 233, 235, 243

control, 82
differences, 234
determinists, 227
factors, 71, 80, 81, 86, 237, 251, 289
filtration, 203, 239
information, 223
stimuli, 84–85
Epiphenomena, 182
Epistemological rationalism, 167
Epistemology, 2, 11, 23, 30, 111, 155, 277, 296–297, 300–301
Cartesian tradition, 166, 285
central problems, 129, 303
empiricist, 17, 24, 163
Hume, 14
of the social sciences, 283
Equality, 93, 99, 149, 218, 220–221, 225, 258, 268
Ethics, 1–2, 260, 264, 266, 303
connection with social science, 253, 273, 275
Ethnicity, 141
Ethnologists, 135
Ethnomethodologists, 31
Euclid, 294
Europe, 37, 139, 147, 153, 157, 177
Evidence, 3, 17, 47, 53, 56, 73, 106, 110, 166
empirical, 56, 65
Evolution
cultural, 237–238, 240–241, 243–245
mechanism of, 83, 203
of cooperation, 211, 215, 225
theory of, 82, 172, 185–186, 195, 206, 209, 212–213, 229
Evolutionary,
biology, 83, 235, 286
environment, 222–223
explanation, 231, 251
game theory (*see* Game theory)
psychologists, 222–225, 227, 231, 237–239
stable strategy, 223
Expected utility. *See* utility
Experimental
conditions, 85, 217
designs, 258
error, 24
method, 118
observations, 17
psychology, 72, 83, 105, 109, 112–113, 275
science, 31, 139
social scientist, 113, 218
subjects, 257
Experiments, 39, 48, 56, 111, 247, 254
on human beings, 7, 235
placebo, 257
quasi, 222, 258
Explanation
causal, 12–13, 42, 46, 51, 53, 56, 125, 127–130
evolutionary, 231, 251
for failure, 19–20
functional, 5, 192–193, 199, 203–205, 236
holistic, 5, 212, 302
increasingly precise, 16
intentional, 111, 118, 120, 161, 173
mechanical, 182
of human action, 37–40, 43, 45, 50, 56–57, 59–60, 64, 68, 72, 122, 176
individuals, 101, 105–106, 141, 169, 181
of human behavior, 21, 71, 75, 175, 237, 265, 297
philosophical, 30
scientific, 15, 31, 54–55, 117, 288, 300
systematic, 12
teleological, 76–78, 202
Explanatory
strategy, 43, 68, 74–77, 92, 118, 144, 191, 199, 265, 277
theory, 39, 174
Extensionality, 111

Fact/value distinction, 277, 281–282, 285, 302
Facts
biological, 180
social, 5, 131–134, 169–182, 186–189, 194–199, 204–205, 248, 273, 278, 283, 302
Fairness, 210, 218, 220–221, 225, 257
Falsifiability, 149
Families, 169, 188, 196, 209
Feminist
empiricists, 285–286
philosophers, 230, 284, 288
philosophers of science, 283–287
theories, 149
Feudalism, 147, 196
Final causation, 236
Fitness, 32, 203, 209, 211, 230, 237–239, 241–245, 249–250

Fitness (*continued*)
- maximizing, 209–210, 212, 216, 218, 220, 223, 226, 228, 251
- reducing, 215, 227, 240

fMRI. *See* Functional magnetic resonance imaging

Fodor, Jerry, 69, 223

Folk biology, 226

Folk physics, 226

Folk psychology, 26–27, 36, 38–40, 42, 53, 56, 69, 80, 83, 86–87, 161–162
- explanatory power of, 60, 68, 77, 84, 144
- limits of, 73
- predictive powers of, 50, 118, 126, 129, 262
- predictive weakness of, 79
- theory of human action, 46, 117, 201
- theory of rational choice, 73, 89, 100–104

Foucault, Michel, 153–158, 180

Free rider, 221, 224, 238

Free will, 2, 35, 117–118, 127, 263–264

Freedom, 140–141, 149, 151, 155, 158, 221, 262, 289, 291
- political, 280
- value, 275, 277, 285

Freud, Sigmund, 38, 143, 152, 157, 160, 163
- psychoanalytic theory, 149, 201
- theory, 142, 144–145, 151–152, 158, 162, 164
- theory of deep meanings, 146, 201

Freudian therapy, 149

Functions, 76, 79, 95, 132, 172, 191–197, 199, 202–207, 210, 222, 236
- biological, 77, 208, 231, 240, 250
- latent, 198, 200–201, 269
- status, 133–134

Functional
- analysis, 82, 194, 198–199, 210, 245
- concepts, 189, 193
- explanation, 79, 193, 203, 236, 304
- kinds, 187
- roles, 198, 243
- traits, 237, 239

Functioning magnetic resonance imaging, 67

Functionalism, 5, 22, 171–173, 191–192, 197, 199–202, 204–208

Functionalist, 199–201, 207, 220

Galileo, 31, 75, 164, 285, 294
- laws, 16, 182

Game theorist, 217–218, 220, 226

Game theory, 213, 226, 251, 267
- argument for cooperation, 220, 237–238
- dominant strategies, 214

Gavagi, 164

Gender, 18, 141–142, 149, 163, 166, 202, 280, 285
- roles, 229–230, 268

Generalization, 14, 20–22, 40, 46, 51–52, 109, 112–113, 127–128, 143, 172, 181–182, 193–194, 199, 226, 243
- biological, 249–250
- causal, 57, 123, 125
- contingent, 55–57
- empirical, 12, 19, 25–26
- exceptionless, 122
- explanatory, 66, 80
- lawlike, 69
- rough and ready, 246
- statistical, 13, 247–249
- sociological, 185–187, 198

Genes, 77, 82–83, 209–210, 212, 220, 222, 229, 234, 235–237, 241, 243–244, 250, 252, 284

Genetic determinism, 232–236, 249

Genetics, 82, 183
- Mendelian, 16, 227, 243
- molecular, 16, 185
- population, 186, 251

Geometry, 15

George Washington, 57–58

Goal, 31–32, 75, 77, 85, 91, 140–142, 208. *See also* Teleology

God, 2, 37, 75–76, 79, 95, 142, 192, 205, 208, 258

Gold, 22

Grammar, 61–62, 119, 121, 159, 225

Greeks, 2, 12, 39

Group mind, 179, 201

Group selection, 211–212, 220–221, 238

Habermas, Jurgen, 157–161, 180, 265

Hamlet, 242

Harvey, William, 75

Hayek, Friedrich, 95

Hegel, Georg Wilhelm Friedrich, 14, 140–141, 153, 157, 159, 171, 176, 295, 304

Hegelian, 149

Heredity, 229, 234, 241
Hermeneutics, 31, 119
Historians, 31, 36, 38, 109, 185
Historicism, 152
History, 39–40, 50–51, 139, 149, 153, 184, 249, 251, 294
  narrative, 249
  of natural science, 75
  of science, 2, 16, 18, 23–24, 31–32, 172, 185, 286
  of scientific progress, 87, 182–183
  philosophy of, 140–142
  revisionist, 252
Hobbes, Thomas, 20, 266, 304
Holism, 5, 170–176, 179–181, 185, 192, 200–202, 302
  and autonomy, 186
  and functionalism, 207–208, 233
  methodological, 187, 191, 197, 199, 210–211
  of the mental, 53–54
Homo erectus, 220, 238
Homo sapiens, 165, 210, 223, 229, 236, 240, 251
Homosexuality, 255
Human
  action, (*see* Action)
    theory of (*see* Folk psychology)
  affairs, 8–9, 11, 94–96, 123, 130, 142–143, 152, 155–156, 158, 192, 195–197, 201, 204–205, 208, 210, 222, 236, 239, 254, 277, 281
  agents, 5, 21, 134, 141, 153, 220
  behavior (*see* Behavior)
  cooperation, 238
  culture, 139, 212, 242–243
  environment, 235
  freedom, 140, 149, 155
  institution, 132, 148, 159, 188, 196
  knowledge, 2
  mind, 223, 225, 237
  populations, 236
  rights, 200, 264
  sciences, 25, 245
  thought, 62
  values, 157
Humanists, 165
Hume, David, 13–15, 46, 128, 304
Humphrey, Laud, 255–256, 269
Huxley, Thomas, 208
Hypotheses, 15, 46, 73–74, 112, 184, 199, 225, 243, 251, 257
  auxiliary, 145
  causal, 125, 134
  explanatory, 14, 66, 170
  mentalistic, 72
Hypothetico-deductivism, 16

Id, 143–145
Ideal gas law, 46, 51, 66, 109, 184
Idealists, 31
Idealizations, 29, 108
Ideals, 148
Identity, 64, 236
Ideology, 148, 156
  critique of, 149, 163, 202
Imperatives, 121, 273
Imperialism, 166
Implications, 11, 37, 113, 117, 146, 185, 270, 287, 294
Incest, 227–228, 230
Indeterminism, 248, 294
Individual selection, 211
Individualism, 5, 301
  methodological, 173, 199, 212, 222
Individualist, 171–174, 199–202, 207, 212, 220–221, 301–302
Initial conditions, 15, 46–47, 49, 51–52, 55, 65–68, 78, 96, 107, 128, 129, 183, 290
Innate, 222–223, 225–226, 235–237
Innatist, 227
Institutions, 90, 94, 146, 154, 156, 159, 173, 179, 202, 207–208, 229, 236, 261, 268, 284, 290
  adaptive, 204, 244
  biological approach to, 246, 249
  creation of, 25, 129, 131–133, 189, 195, 269
  meanings of, 28, 130, 148, 162, 202, 278, 291
  political, 73, 267, 280
  social, 26, 79, 95, 114, 117, 130, 141, 144, 147–148, 153, 158, 162, 169, 178, 188, 204–205, 227, 258, 297
    creation of, 36, 134–135, 210–212, 222, 289
    identification of, 198–201
Instrumentalism, 105, 109, 301–302
Instrumentalist, 106, 108–111, 114–115
Intelligent designer, 75

Intelligibility, 5, 25, 38, 43, 45, 111, 120, 129, 152, 163–164, 166–167, 262, 265, 279, 295–297
Intensionality, 63, 65, 67, 69, 111, 184
Intentional states, 63–64, 72, 76–78, 117, 131–132, 142, 266, 270, 300
Intentionality, 62–65, 86–87, 117, 130, 149, 161, 185, 264, 298
  collective, 131–132
  problems, 69, 73, 111, 113, 170, 192, 300
Interpretation, 43, 73, 114, 119, 139, 141, 153, 165, 251, 277–279
  and prediction, 31, 253, 296
  causal, 118
  instrumentalist, 106, 110
  of human behavior, 25–28, 90, 100, 104, 166
Interpretationalism, 140, 142, 176, 301
Interpretationalist, 38, 60, 65, 117–118, 122, 125, 130, 134, 143, 152, 248, 287–288, 298–299, 304
Introspection, 73, 78, 97, 166
IQ, 230, 234–235, 289

Justice, 1–3, 31, 35, 140, 149, 210, 256, 258, 262, 291
  Rawls's theory of, 267–268, 270, 281–282
Justification, 2, 11, 106, 108, 208, 262, 267, 279, 295–297, 302
  moral, 257, 259

Kant, Immanuel, 158, 260, 263–266, 304
Kepler, Johannes, 75, 285
  laws, 16, 182
Kin relatedness, 230
Kinetic theory of gases, 51, 109–110
Knowledge, 12, 23, 135, 145–146, 151–152, 157–158, 238, 253, 273–274, 295–297
  causal, 13–14, 46, 118, 125–126, 163
  moral, 254, 275, 296
  normative, 160, 303
  objective, 18, 131, 285, 288
  predictive, 5–6, 8, 158, 262
  requirements of, 175–176
  scientific, 3, 15–16, 19, 24, 30, 53, 56, 68, 111, 155, 182, 264, 278
  social scientific, 257, 261, 290–291
  theory of, 2, 11, 165, 265, 277 (*see also* Epistemology)
Kuhn, Thomas, 23–24, 277, 294

[L], 39–43, 45–52, 56–59, 65–69, 74, 79–80, 89–91, 100, 143, 161, 170, 265
  and [LE], 84, 86
  as a causal law, 42
  as unfalsifiable, 53
  function of, 125
  improved, 41, 46, 184
  methodological problems of, 176
  predictive content of, 50
  truth of, 104, 126, 166
Language, 16, 24, 61, 64, 67, 95, 130–131, 159, 161, 164–166, 173, 183, 187, 230, 242, 302
  learning, 65, 119–120, 125–126, 132, 225
  ordinary, 134, 197, 235, 277
Latent functions, 198, 200–201, 269
Lavoisier, Antoine, 23, 285
Law of effect, 80–81, 84–85, 250
Lawlike generalizations, 69
Laws
  and causation, 13, 28–29, 43, 51, 56, 117, 122, 127–128
  and explanation, 77, 186
  and prediction, 31, 79, 142, 298
  contingency of, 56, 59
  general, 18, 172, 251
  in biology, 249
  of human behavior, 19–21, 88, 115
  of nature, 15
  providing knowledge, 23
  physical, 57, 78, 145, 183, 274
  psychological, 170, 172, 181, 183–184, 194
  See also Causal: laws; Galileo: laws; Kepler: laws; Mendel: laws; Newton: laws; Psychology: laws
[LE]. *See* Law of effect
Legal system, 256, 259
León, Ponce de, 64
Lévi-Strauss, Claude, 155, 192
Life science, 77, 172, 236
Linguistic meaning, 120, 132
Literary criticism, 119
Locke, John, 266
Logic, 1–2, 166, 224, 295
Logical
  argument, 45
  connection, 39, 45, 55–59, 68, 127–128
  deduction, 15
  empiricism, 14 (*see also* Logical positivism)

incompatibility, 163
positivist, 24, 71, 152, 179, 182
property, 63
Logical positivism, 14, 18, 23, 140–141, 150, 172, 295
Lois Lane, 62–63

Machiavelli, Niccolò, 20, 33
Macrosocial
facts, 301
laws, 170
phenomenon, 170, 197
science, 187, 194
Maladaptation, 197, 210, 230, 240, 249
Male domination, 18, 287
Marginal utility. *See* Utility
Marginalism, 98
Marginalist, 90–91, 93, 95–98
Margins for error, 51
Marx, Karl, 5, 142–143, 146–148, 154–157, 304
Marxian
interpretations, 137, 201–202
sociologist, 198
theory, 22, 141, 149, 151–153, 162–164, 200, 206, 289
Marxism, 142, 146, 149–151, 163
Materialism, 294, 299–300
Mathematical
economists, 94, 98, 302
models, 69, 107, 248, 251
sociologists, 248
truths, 38, 53
Mathematics, 2, 39, 53, 230
Meaning, 5, 31–32, 50, 60, 122
cultural, 30
deep (*see* Deep meanings)
nature of, 166
of actions, 30–32, 53, 135, 160–162, 166, 173–174, 201, 295 (*see also* Action)
of concepts, 39
of history, 140, 149
of institutions, 28, 130, 148, 162, 202, 278, 291
of words, 16, 95, 120, 132, 164, 235
theory of, 16
Means of production, 147–148, 151
Measurability, 68, 97–98
Mechanism, 32, 72, 82–83, 193, 221, 230, 236, 248, 294
causal, 15, 24–25, 76–77, 123, 127, 142, 250, 304
of evolution, 203–204, 209, 228
of heredity, 241
Mechanistic theory, 76
Memes, 242–245, 250
Mendelian
genetics, 16, 227
laws, 243
ratios, 242
Mental
health, 121, 144, 153–154, 229–230, 234, 262
images, 242
mechanisms, 72
module, 223, 225, 239
processes, 67
states, 68, 71–75, 105, 113, 131, 180, 184, 247, 298
Mental-state description, 67
Metaphysics, 2, 29, 78, 153, 295, 297, 299–303
Methodological. *See* Holism.
Methodological individualism. *See* Individualism.
Methodology, 8, 11, 72, 192, 288
economic, 161
empirical, 13, 176
of social science, 43, 259
statistical, 13
Milgram, Stanley, 255
Mill, John Stuart, 260, 267, 304
Mind, 48, 61, 72–74, 98, 114, 123, 131, 141–142, 176, 202, 223, 225–226, 238, 242–244, 266
philosophy of, 301
relation to body, 55, 68, 111, 298–299
relation to brain, 2, 222, 237, 299
See also Group mind
Mind-body problem, 67–68
Module, 223, 225–226
hardwired, 237–239
Molecular biology. *See* Biology: molecular.
Moral
agents, 161, 264
conflicts, 259, 261, 263, 265, 270
decision, 161, 261, 265, 267, 270
knowledge, 30, 158, 254, 275, 296
philosophy, 5, 30, 260, 263, 270, 279, 302–303

Moral (*continued*)
- problems, 254, 256, 258–259, 303
- questions, 256, 271, 273–275, 280
- rules, 258, 266–267

Morality, 19, 200, 264–265, 267
Morgenstern, Oskar, 102
Mother nature, 237, 288
Motives, 158, 166, 226, 233, 264, 270
Multiple realizability, 188–189, 194, 197
Mutation, 77, 211, 221, 234, 243–245

Nation-state, 147
Nativism, 223, 227, 230–231, 237
Natural kinds, 22, 27, 30, 79, 184–185
Natural rights, 265–266, 271
Natural science
- aim of, 298
- explanations, 12, 42, 75, 279
- history of, 31
- methods of, 4, 7, 11, 28, 31, 159, 165, 176, 299
- progress in, 7, 21, 23, 25, 273, 294–295

Natural selection, 5, 76, 82, 204–205, 218–221, 225, 229, 231, 236–237, 241–242, 249, 294. *See also* Evolution: theory of
- and blind variation, 21, 77
- cultural, 208–209, 243–245
- mechanism of, 203, 210–212, 239, 250

Naturalism, 131, 140, 287–288, 299–300
- aims, 266
- opponents, 27, 31, 279, 298
- problems, 65, 111, 135, 176

Naturalistic social science, 57, 111, 134–135, 142, 287, 304
Nature versus nurture, 207, 235
Necessary and sufficient conditions, 175
Neologism, 22, 277
Neuroeconomics, 67
Neuroscience, 2, 67–68, 79–80, 83, 87, 96, 113, 127, 142–143, 145, 172, 184–187, 300
Neurosis, 143–144, 146, 151, 158, 163
Neutrality, 161, 254, 273–276, 288, 302–303
Newton, Isaac, 31, 75–76, 80, 182, 285, 294
- laws, 13, 16, 52, 109, 145, 182
- mechanics, 183, 299
- physics, 25, 300
- revolution, 20
- theory, 24

Niche construction, 238
Norm of reaction, 235–236
Normativity, 161
Norms, 118–121, 125, 132–135, 141, 143–144, 153–155, 195, 197, 208, 235, 238
- cultural, 210
- moral, 205, 242, 274
- of conduct, 178
- social, 123–124, 133, 147–148, 178, 195, 212, 227, 229, 231

Objectivity, 18, 134–135, 267, 274, 276–277, 284, 287–288
Observation, 13, 20, 22, 39, 66, 78, 114, 118, 164, 174, 247, 255–256, 258–259, 265, 283–284, 286, 301
- and prediction, 15–18, 24, 105–106, 134, 175, 254

Oedipus complex, 162
Olds, James, 87
On the Origin of Species (Darwin), 208–209
Operant
- behavior, 80–87, 115
- conditioning, 22, 82
- response, 85
- theory, 84

Opportunity costs, 91
Organicism, 171, 208
Ornithologist, 229
Other minds, 73, 226

Paradigms, 24–25, 77, 178, 208, 298
Pareto optimal, 99, 269, 276, 302
Pareto, Vilfredo, 162, 261
Pavlov, Ivan, 80
Phaedo (Plato), 119
Phenomenology, 166
Phenotypic
- outcomes, 234–235
- traits, 236, 243
- variation, 236

Philosopher of biology, 237
Philosopher of science, 289
- feminist, 287

Philosopher of social science, 68, 149, 152, 187, 245, 274
- feminist, 287–288

Philosophy of history, 140–142
Philosophy of language, 134, 166

Philosophy of mind, 301
Philosophy of psychology, 59, 62, 69
Philosophy of science, 3, 14, 18, 33, 46, 72, 88, 105, 152, 260, 285, 288, 294–295, 301
  empiricist, 17, 33, 157
  feminist, 284, 287
Philosophy of social science, 4, 6–7, 9, 21, 29, 139–140, 146, 159, 208, 278
  interpretative, 297–298
  problems, 46, 94, 144, 246, 293, 295, 303–304
Phlogiston, 23, 300
Photons, 86, 171
Physical sciences, 76, 186–188, 246–247, 254, 301
Physics
  development, 294
  facts, 174–175
  folk, 226
  history of, 24
  knowledge, 8, 20
  laws, 25, 78, 145, 248
  models, 29, 51
  quantum, 25
  theories, 75, 145, 172, 185, 274
  See also Newton: physics
Physiology, 35, 67, 82, 179, 183, 203, 226, 286
Plato, 29, 118, 200, 293, 295, 304
Pleasure center, 87
Pleasure principle, 143
Pluralism, 280, 287
Political philosophy, 1–2, 267, 302
Political science, 2, 18, 39–40, 71, 73, 88–89, 94, 280
Population biologist, 236
Population genetics, 186, 251
Populations, 186, 195, 209, 236, 243, 245, 251
Positivism. *See* Logical positivism
Postmodernism, 31
Poverty of stimulus argument, 225, 238
Prediction
  and explanation, 15, 52, 72, 87, 101, 105–106, 254, 274
  and interpretation, 31, 253
  improvements in, 6, 112, 125, 163, 246, 295–296
  novel, 248
  observational, 17, 30, 105, 175
  power of, 16
  precise, 12, 50–51, 126, 146
  success of, 15, 46, 164
Predictive knowledge, 5–6, 8, 125–126, 158, 262, 296
Prescriptions, 23, 142, 200, 254, 261, 277, 302
Principle of charity, 165–166
Prisoner's dilemma, 213, 215–216, 267
  iterated, 217–218
Privacy, 256, 258, 261–262, 264, 291
Probability, 78, 103, 199, 216
Problem of composition, 194
Progress
  as reduction of previous theories, 182–183
  economic, 289
  history of, 16, 87, 185
  in advancing knowledge, 11
  in moral philosophy, 279
  in science, 7, 278–279, 287, 295
  in social science, 7, 23, 25, 27, 38, 273
  lack of, 22, 26, 46, 273
  opponents of, 24
  rate of, 20
  technological, 28
Proletariat, 149, 200
Propositional content, 60–62, 145
Pseudoscience, 27, 152, 162, 186, 278
Psychoanalysis, 22, 142, 145–146, 162, 201
Psychological
  factors, 169–170, 177, 180–182, 237
  facts, 170, 178, 186, 189, 194
Psychologists, 1, 36, 38, 62, 73–74, 79, 81, 86–87, 111, 113, 169, 172, 220, 267, 294
  evolutionary, 222–225, 227, 238–239
  experimental, 6, 109, 275
  intentional, 184–186
Psychology
  as the study of behavior, 74, 78, 80
  cognitive, 68–69, 113
  cognitive social, 112
  evolutionary, 222, 231
  experimental, 72, 83, 105, 112–113
  folk (*see* Folk psychology)
  individual, 172
  intentional, 183–185
  philosophy of, 59, 62, 68–69, 187

Psychology (*continued*)
  scientific, 74, 184
  social, 280–281
Psychopathology, 87, 143, 151
Psychosis, 143
Psychotherapy, 144–145
Public goods, 251
Pure reason, 263, 265
Purpose, 31–32, 37, 74–79, 85, 131, 140–142, 201, 208, 250. *See also* Teleology
PV=nRT. *See* Ideal gas law

Quantum mechanics, 16, 24, 78, 183
Quarks, 19, 78, 171
Quasi-experiments, 222, 258
Quasi-organisms, 208
Quine, Willard Van Orman, 70, 164–166

Race, 18, 141–142, 163, 166, 202, 280, 287
Racism, 27, 230–231
Random variation, 203, 209, 211, 240–241
Rational choice theory, 20, 30, 42, 51, 69, 73, 89–92, 94–95, 112, 161, 212–213, 215, 266–270
  and instrumentalism, 106, 109–110
  falsity of, 99
  model, 29, 104, 251
  predictions, 100, 106, 113–114
Rationality, 43, 90, 98, 100, 105, 107, 113, 159, 161–162, 214, 267, 277
Rawls, John, 267–269, 282
  theory of justice, 267, 269–270, 281–282
Reality principle, 143
Reasons
  and causes, 43–45
  explanatory power of, 60
  for action, 43, 60, 134
  for suicide, 177–178
  laws about, 60
Reciprocal altruism, 230
Rectilinear motion, 80
Reduction, 16, 24, 80, 145, 172, 182–185, 203, 287, 301
Reflexivity, 159
Regularities, 12–13, 22, 30–31, 60, 117, 123, 170, 177, 179, 182–183, 187, 194, 197, 226, 250–251
  biological, 251
  causal, 19, 119, 128
  statistical, 36, 106, 247–248
Reinforcement, 81–86, 220
Relativity, 16, 24, 49, 131, 183, 194
Representation, 61–63, 84–85, 212–223, 283, 300
Repression, 22, 143–144, 149, 255
Revealed preference, 100–102, 104, 276
Revisionist, 252, 288
Richerson, Peter, 244–245
Rights, 2, 133, 147, 198, 213, 256, 263, 268–270, 278, 281, 303
  individual, 200, 254, 259–260, 264, 291
  natural, 265–266
Roles, 16, 129–130, 146–147, 154, 173, 202
  explanatory, 44
  functional, 198, 243
  gender, 229–230
  social, 129, 133–134, 158
Rousseau, Jean-Jacques, 219, 266
Rule *R*, 122–124
Rules, 122–123, 125–126, 129–135, 154, 161, 197, 268, 278
  learning, 220, 238, 242, 245
  linguistic, 119–121, 225
  marriage, 36, 121, 130, 147, 194
  moral, 258, 266–267
  private, 124
  social, 27, 124, 147, 173, 224

Same cause, same effect, 128, 177–178
Science
  aim of, 43
  autonomous, 172
  empirical, 4, 157, 165, 254, 265
  fallibility of, 16
  history of, 2, 16, 18, 23–24, 31–32, 172, 185, 286
  methods of, 14, 286
  progress (*see* Progress: in science)
Scientific explanation, 15, 31, 37, 42, 47, 117, 288, 300
Scientific knowledge, 3, 15, 18–19, 24, 53, 68, 111, 155, 158, 182, 264, 278, 296–297
Scientific progress, 87, 182–183, 278–279, 287, 295
Scientific psychology, 74, 184
Scientific standards, 25, 68
Scientism, 152
Searle, John, 130–133, 154, 159
Selection
  cultural, 237, 245

Darwinian (*see* Evolution; Natural Selection)
group, 211–212, 220–221, 238
individual, 211–212
Self-interested agent, 218, 266
Sen, Amartya, 279–282
Sensation, 74, 86, 300
Sexual orientation, 141–142, 163, 283
Skepticism, 72, 164, 166, 171, 242, 285
Skinner, Burrhus Frederic, 12, 72, 74, 80–82, 144, 250, 262, 300
Smith, Adam, 12, 92–94, 114, 269
Social
class, 5, 170, 194, 202, 283
contract theory, 266–267
facts, 5, 131–134, 169–182, 186–189, 194–199, 204–205, 248, 273, 278, 283, 302
forces, 18, 141, 169, 178, 180, 182, 197
groups, 124, 134, 170, 210
institutions, 26, 79, 95, 114, 117, 130, 141, 144, 147–148, 153, 158, 162, 169, 178, 188, 204–205, 227, 258, 297
creation of, 36, 134–135, 210–212, 222, 289
identification of, 198–201
integration, 178, 181–182, 198–199
knowledge, 19, 160, 273
life, 21, 27, 144, 153, 156, 158, 195, 244
norms, 123–124, 133, 147–148, 178, 195, 212, 227, 229, 231
organization, 147, 150, 200
phenomena, 36, 73, 134–135, 186, 200, 207, 209, 234, 244, 247, 251, 301
practices, 195, 207, 219
processes, 19, 90, 142, 146, 149, 156, 197, 244, 273, 291
role, 129, 133–134, 158
rules, 27, 124, 147, 173, 224
Social science
aim, 26–27, 64, 120, 135, 163, 291, 297
as a young science, 19, 23
as naturalistic, 57, 111, 134–135, 142, 287, 304
empirical, 69, 251, 266
failure of, 19, 53
goals, 129, 254, 288, 291, 295
methods of, 119, 126
naturalistic, 57, 111, 134–135, 142, 287, 304
problems, 7, 134, 258, 293
progress, 7, 23, 25, 27, 38, 273
Socialism, 142, 147–148, 150, 157
Society, 20, 27–31, 96, 126, 139–140, 144, 147–152, 154–157, 160, 165, 173–175, 181, 192–193, 203–207, 210–213, 227, 229, 230, 240, 258–259, 261, 266–269, 276, 288–289, 301
as a whole, 170–172, 178–179, 198–202, 255, 290–291
capitalist, 30, 149–150
construction of, 129–131, 133–135,
Sociobiology, 22, 173, 201, 222, 231, 289
Sociologists, 1, 6, 18, 36, 73, 109, 123, 129–130, 157, 169, 181, 186, 220, 248
Sociology
autonomy of, 172, 179, 181, 183, 185–186, 194
empirical, 180
modern, 153
methods of, 136, 176
of science, 18, 277
Soft sciences, 297
Soviet Union, 94, 140, 149
Speech, 25, 40, 48–49, 82, 161
acts, 125, 132–133, 154, 159
Spirit, 141, 176, 202
Stag hunt, 219–220
Standpoint theory, 283–285
State of belief, 85
Statistical regularities, 36, 106, 247–248
Statistics, 177, 246
Stimulus, 80–81, 84–86, 225, 238
Structuralists, 31
Structure of Scientific Revolutions (Kuhn), 23
Subcultures, 135, 156, 275
Subjectivity, 63, 131
Substructure, 147, 151, 155–157
Superstructure, 147–148, 153–155, 157, 202
Sucker, 267
Suicide, 25, 177–182, 198–199, 203, 248
Superego, 143–145
Superman, 62–63
Supervenience, 188–189
Supply and demand, 91–92, 98
Synapses, 61–62, 79, 299

Tastes, 100, 102, 156, 268, 285
Teleology, 76–79, 82, 95, 172, 203

Testability, 66, 86, 145, 149, 174
Theology, 119
Theory of duties and rights, 260, 263
Theory of knowledge. *See* Epistemology
Theory of other minds, 226
Theory of revealed preference, 100–102, 104
Therapeutic prescription, 144–145
Thermodynamics, 109–110, 170, 183–184
Thucydides, 12, 20
Tit for Tat, 216, 218–219, 238
Titanic, 57–59, 62
Token, 81, 132
Tolerance, 275, 296–297
Totalitarianism, 200–201, 275
Traits, 208, 211, 227, 231, 236, 238, 242–243
  adaptational, 77, 230, 239
  biological, 195, 250, 252
  cultural, 240–241
  genetic, 221, 223, 237, 251
  group, 210, 213
  hardwired, 204, 223
  heritable, 76, 203, 209–210, 229, 235, 249
  socially significant, 233–235
Transitivity, 98, 100, 102
Translation, 24, 125–126, 164–165, 174–175, 179
Truth, 26, 39, 52, 62–63, 105, 107–108, 110, 126, 135, 142, 154, 165, 175, 180, 248, 252, 258, 265, 286–287
  a priori, 166, 263, 300
  and falsity, 16, 104, 184
  causal, 246
  contingent, 55
  mathematical, 38, 53
  objective, 24
Type, 8, 25, 42, 62, 82, 160, 180, 187–189, 197, 199, 223, 247

Ultimatum game, 218, 220
Underdetermination, 17
Undetectable mechanisms, 72
Unity of science, 179
Untestability, 145
Utilitarianism, 260–261, 264, 267, 269, 281–282
Utility, 260, 268
  cardinal, 91, 96–97, 101, 162, 302
  expected, 102–104
  marginal, 91, 95–97
  maximizing, 91, 93, 103, 112, 214–216, 226, 261, 277, 281
  ordinal, 98–101, 106–107, 161–162, 269, 276
  theory of, 90

Vagueness, 42, 49
Value ladenness, 274–275, 277–278, 283
Value neutrality, 161, 275–276, 302
Values, 27, 51, 66, 103, 109–110, 140, 144, 146, 148, 153, 165, 182–183, 274–277, 281, 283–288, 302
  cultural, 156, 268
  human, 157, 239, 278, 285
  measurement of, 52
  moral, 273, 279
  See also Fact/value distinction
Verification, principle of, 16
Voltaire, 195
Voluntary, 39, 81, 133, 148, 180
Von Neumann, John, 102

Wason selection test, 224
Weber, Max, 28, 37–38, 42–43, 45, 153, 304
Westermack, Evart, 228

Xenophobia, 27, 230–231, 234