Fourth Edition

Alex
Rosenberg
and Lee
McIntyre

PHILOSOPHY
OF SCIENCE

A Contemporary
Introduction

Philosophy of Science

Any serious student attempting to better understand the nature, methods, and justification of science will value Alex Rosenberg and Lee McIntyre's updated and substantially revised fourth edition of *Philosophy of Science: A Contemporary Introduction*. Weaving lucid explanations with clear analyses, the volume is a much-used, thematically oriented introduction to the field.

The fourth edition has been thoroughly rewritten based on instructor and student feedback, to improve readability and accessibility, without sacrificing depth. It retains, however, all of the logically structured, extensive coverage of earlier editions, which a review in the journal *Teaching Philosophy* called "the industry standard" and "essential reading."

Key Features of the Fourth Edition:

- Revised and rewritten for readability based on feedback from student and instructor surveys.
- Updated text on the problem of underdetermination, social science, and the realism/antirealism debate.
- Improved continuity between chapters.
- Revised and updated Study Questions and annotated Suggested Readings at the end of each chapter.
- Updated Bibliography.

For a list of relevant online primary sources, please visit: www.routledge.com/9781138331518.

Alex Rosenberg is R. Taylor Cole Professor and Chair in the Philosophy Department at Duke University. He is also co-director of Duke's Center for Philosophy of Biology. Rosenberg has held fellowships from the National Science Foundation, the American Council of Learned Societies, and the John Simon Guggenheim Foundation. In 1993, Rosenberg received the Lakatos Award in the philosophy of science.

Lee McIntyre is a Research Fellow at the Center for Philosophy and History of Science at Boston University. He is the author of *Respecting Truth* (2015); *Post-Truth* (2018); and *The Scientific Attitude* (2019).

This innovative, well-structured series is for students who have already done an introductory course in philosophy. Each book introduces a core general subject in contemporary philosophy and offers students an accessible but substantial transition from introductory to higher-level college work in that subject. The series is accessible to non-specialists and each book clearly motivates and expounds the problems and positions introduced. An orientating chapter briefly introduces its topic and reminds readers of any crucial material they need to have retained from a typical introductory course. Considerable attention is given to explaining the central philosophical problems of a subject and the main competing solutions and arguments for those solutions. The primary aim is to educate students in the main problems, positions and arguments of contemporary philosophy rather than to convince students of a single position.

Recently Published Volumes:

Virtue Ethics
Liezl van Zyl

Philosophy of Language
3rd Edition
William G. Lycan

Philosophy of Mind
4th Edition
John Heil

Philosophy of Science
4th Edition
Alex Rosenberg and Lee McIntyre

For a full list of published Routledge Contemporary Introductions to Philosophy, please visit www.routledge.com/Routledge-Contemporary-Introductions-to-Philosophy/book-series/SE0111.

Philosophy of Science

A Contemporary Introduction

Fourth Edition

Alex Rosenberg and Lee McIntyre

Contents

13 Naturalism in the Philosophy of Science   **223**

*Naturalistic models purport to rescue the understanding of scientific
     explanations from earlier worries about induction, rationality, and
     underdetermination, by grounding philosophy of science within science
     itself. If the understanding of science is merely of a piece with science
     itself, might its justification rest on firmer ground? What are the costs?*

14 The Contested Character of Science   **237**

*One final challenge to the rationality and logic of science comes from
     critiques based on postmodernism, scientism, and sexism. Sociological
     issues about the practices of science provide a counterbalance to the
     traditional, logical understanding of scientific explanation as distinctive
     from other human activities.*

15 Science, Relativism, and Objectivity   **256**

*This book concludes with an exploration of relativism and Kuhn's problem
     of the alleged "incommensurability" of some scientific accounts
     with others. Is there a sense in which those who are working within
     different conceptual schemes are explaining different worlds?*

Preface to the Fourth Edition

By the time a fourth edition of an RCIP volume is needed, it may become difficult to make further revisions and improvements, beyond required updates to the annotated Suggested Readings at the end of each chapter and the book's General Bibliography. This important material has indeed been made *au courant* here, but the fourth edition is also thoroughly different in dozens of other unobvious, and some detectable, ways.

Readers will notice that many sections of the book have been revised and rewritten to increase readability, following the suggestions made by numerous instructors who used the third edition. These same readers, with a concern for cohesion and flow in the volume, will appreciate the improved continuity between chapters. In addition, the problem of underdetermination, social science, and the realism/antirealism debate are important issues in the text that have been brought up to date in this new edition. Finally, many instructors who make use of the Study Questions at the end of each chapter will notice that they have been revised and augmented.

As with the first three editions, Alex still retains the aspiration of approaching that great classic of philosophy and pedagogy, Carl G. Hempel's *Philosophy of Natural Science*. A splendid example is a great spur. "Ah, but a man's reach should exceed his grasp," said Robert Browning. He might as well have been talking about the first, second, and third editions of this book. It remains true about the fourth. The aspiration of Lee McIntyre, who has been brought on as a co-author for the fourth edition, has been to make the entire work more accessible than the previous editions, as well as to supplement and modify Alex's take in a few key areas.

As with the earlier editions Alex remains eager to show that the problems of the philosophy of science are among the most fundamental issues of philosophy, mere substitution instances of questions on the agenda of the discipline since Plato.

The new edition preserves one of the key innovations in the third edition, which was to key the Suggested Readings to articles reprinted in the three most useful anthologies in the business currently available, Martin Curd and J. A. Cover's *Philosophy of Science: The Central Issues*, Marc Lange's *Philosophy of*

Science: An Anthology, and Yuri Balashov and Alex Rosenberg's *Philosophy of Science: Contemporary Readings*. Besides broader suggestions, each chapter's Suggested Readings section ends with specific recommendations from some or all of these three books. Alex's own course in the philosophy of science always focuses on the canonical papers that these collections anthologize. Assigning chapters from this book enables a reduction of time devoted to lectures that merely set the stage for a close study of the real achievements of our discipline. Alex prepared a website posted by Taylor & Francis (www.routledge.com/9781138331518) of important open access and JSTOR papers that can be combined with the third edition. The same list will serve this function for the fourth edition.

1 The Relationship between Philosophy and Science

Overview

Philosophy of science is a difficult subject to define, in large part because philosophy is difficult to define. But for at least one controversial definition of philosophy, the relations between the sciences—physical, biological, social, and behavioral—and philosophy are so close that philosophy of science must be a central concern of both philosophers and scientists. On this definition, philosophy deals initially with questions that the sciences cannot yet or perhaps can never answer, and with the further question of why the sciences cannot answer these questions.

This chapter argues for the adequacy of this definition in a number of different ways. It shows how the sciences emerged successively from philosophy, how the subdivisions of philosophy are related to the sciences, and how the history of philosophy reflects an agenda of problems set by the sciences.

What Is Philosophy?

Philosophy is not an easy subject to define. Its etymology is obvious—the love of wisdom—but unhelpful to someone who wishes to understand what the discipline of philosophy is about. Nor is it enough to know what the most important sub-disciplines of philosophy are. 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 and why they constitute one discipline (philosophy), instead of being parts of other subjects, or their own independent areas of inquiry.

The major sub-disciplines 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**[1] 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 there are 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 non-physical?" or "Do I have free will?" are all metaphysical questions that most people have asked themselves.

But knowing these four domains of inquiry may just deepen the mystery of what philosophy is. They 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 priests, ministers, imams and others who deliver sermons? Whether we have free will, or if 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. Yet, none of these disciplines or approaches in fact explores any of these questions in the way that philosophers pursue them. The problem thus remains, what makes all of them parts of a single discipline, philosophy?

What is worse, we now have another question that will certainly occur to the reader of this book. The one compartment of philosophy that was not even mentioned in the list of its chief sub-disciplines is the philosophy of science. Yet that is the subject of the very book in your hands. Where does it fit in and how important can it be if it is not one of the four main areas of philosophical inquiry?

One answer to the question of what philosophy is makes the philosophy of science at least as central to the whole subject as logic, ethics, epistemology, and metaphysics. It also resolves the other matter of what makes one discipline out of these diverse topics. Nevertheless, the definition of philosophy to be offered below is tendentious. It is a partisan definition, reflecting a distinctive point of view. In deciding whether you want to accept it, ask yourself whether other

definitions can synthesize the diverse questions philosophers address better than this one:

Philosophy deals with two sets of questions:

First, questions that science—physical, biological, social, behavioral—cannot answer now and perhaps may never be able to answer.

Second, questions about why the sciences cannot answer the first set of questions.

Philosophy and the Emergence of the Sciences

There is a powerful argument for this definition of philosophy in terms of its historical relationship with science.

Technology and engineering began in many places independently, and advanced more rapidly in some places than others. China is the source of many of the most important advances in technology—paper, printing, gunpowder, and probably the magnetic compass, to name only the most obvious. Science, however, seems to have begun in the Near East, and to have taken off among the Greeks.

The history of science from the ancient Greeks to the present is the history of one compartment of philosophy after another breaking away and emerging as a separate discipline. Each of these disciplines that have spun off from philosophy, however, has left to philosophy a set of distinctive problems: issues they cannot resolve, so must leave either temporarily or permanently for philosophy to deal with. Thus, by the third century BC, Euclid's work had made geometry a "science of space" separate from but still taught by philosophers in Plato's Academy.

Soon after, Archimedes was calculating the approximate value of the irrational number π and finding ways to calculate the sum of an infinite series. But almost from the outset of its history as a discipline distinct from philosophy, mathematics turned its back on a series of questions that one might have thought would interest mathematicians profoundly.

Mathematics deals with numbers, but it cannot answer the question what a number is. Note that this is not the question what "2" or "dos" or "II" or "$10_{(base\ 2)}$" is. Each of these is a numeral, an inscription, a bit of writing, and they all name the same thing: the number 2. When we ask what a number is, our question is not about the symbol (written or spoken), but apparently about the thing. Philosophers have been offering different answers to these kinds of questions at least since Plato held that numbers were particular things—albeit abstract things not located in space and time. By contrast with Plato, other philosophers have held that mathematical truths are not about abstract entities and relations between them, but are made true by facts about concrete things in the universe, and reflect the uses to which we put mathematical expressions. Yet 2,500 years after Plato lived, there is still no general agreement on the right answer to the question of what numbers are.

The work of Galileo and Kepler, and Newton's revolution in the seventeenth century, made physics a subject separate from metaphysics. To this day, the name of some departments in which physics is studied is "natural philosophy." But physics too has left profound problems to philosophy for centuries. Here is an important instance.

Newton's second law tells us that F = ma, force equals the product of mass and acceleration. Acceleration in turn is dv/dt, the first derivative of velocity with respect to time. But what is time? Here is a concept we all think we understand, and one that physics requires. Yet both ordinary folk and physicists, for whom the concept is indispensable, would be hard-pressed to tell us what exactly time is. Notice that to define time in terms of hours, minutes, and seconds is to mistake the units of time for what they measure. It would be like defining space in terms of meters or yards. Space is measured with equal accuracy in meters or yards. But suppose we ask which is the correct way of measuring space? The answer of course is that there is no uniquely correct set of units for measuring space; yards and meters do equally well. By the same token, neither can be said to "define" or constitute space. The same goes for time. Seconds, centuries, millennia are just different amounts of the same "thing." And it's that thing, time, which comes in different amounts that we want a definition of. We could say that time is duration, but then duration is just the passage of time. Our definition would presuppose the very notion we set out to define.

Explaining exactly what "time" means is a problem that science left to philosophy for a period of at least 300 years. With the advent of the special and general theory of relativity, physicists claimed a share in trying to answer this question again. Albert Einstein's own reflections on time, which led to the conclusion that time intervals differ between different frames of reference—points from which the durations are measured—owes much to philosophers' critique of Newton's conception of absolute space and time as independent containers in which things can be absolutely located and dated. Even today, while several important physicists address the question of why time has a direction, none take on the question of what time itself is. The matter is either premature or beyond physics.

Until the end of the nineteenth century, many chemists treated the question of whether there were atoms or not as beyond the reach of their discipline. Their refusal to debate the question stemmed from their theory of knowledge. The winner of the debate about whether atoms exist, Ludwig Boltzmann, one of the greatest scientists of the era, went to his death believing he had lost the epistemological argument that we can have knowledge about atoms.

In biology the shift of questions from philosophy's side of the ledger to science's is particularly clear. It was only in 1859 that *The Origin of Species* finally set biology apart from philosophy (and theology). Many biologists and not a few philosophers have held that after Darwin, evolutionary biology took back from philosophy the problems of explaining human nature or identifying the purpose or meaning of life. These biologists and philosophers hold that Darwinism shows that our nature differs only by degrees from that of other

animals. They argue that Darwin's great achievement was to show that there is no such thing as purposes, goals, ends, meaning, or intelligibility in the universe; that its appearance is just an "overlay" we confer on the adaptations we discern in nature. Adaptations are really just the result of the environment's persistent filtration of blind variations creating the appearance of design. It is for this reason that evolutionary theory is so widely resisted. Some people reject the answers that biology gives to questions about purpose, meaning, and human nature. Instead they turn to philosophy or religion. Whether one agrees with Darwin's theory of natural selection or not, it is an impressive example of how scientific research leaves some questions to philosophy for centuries, and then takes them on when it finally considers itself equipped to do so.

In the last century psychology broke free from philosophy as a separate discipline, and began to address questions about the nature of the mind, the self, the will, and consciousness, which philosophy had been taking seriously for two and a half millennia. And of course, in the last 50 years, philosophy's enduring concern with logic has given rise to computer science as a separate discipline.

The lesson is clear. Every science is the child of philosophy. Each eventually moves out, but ends up leaving "baggage" at home.

Science and the Divisions of Philosophy

There are other questions that 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. Scientists may have views on these matters—and in fact may disagree with one another as much as non-scientists do—but since scientists generally agree on the broadest matters of their sciences, it is hard to avoid the conclusion that science can't decide these questions.

Questions about what ought to be the case, what we should do, what is good and bad, right and wrong, just and unjust, are called "**normative**." By contrast questions in science that are presumably descriptive are said to be "positive." Many normative questions have close cousins in the sciences. Thus, psychology will interest itself in why individuals hold some actions to be right and others wrong, anthropology will consider the source of differences among cultures about what is good and bad, political scientists 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—whether social or natural—do not challenge or defend the normative views that we may hold.

This raises two questions: First, whether or not science by itself can address questions like "Is it permissible to destroy embryos for stem cell research?" And if science cannot decide this matter, the question arises "Why can't science answer this question?" Notice that both of these questions are addressed in philosophy. They are instances of the two kinds of questions by which we have defined the discipline. Of course at various times, including the present, some philosophers and scientists have tried to show that science can in fact answer at least some

if not all normative questions. If it can do so, we will have eliminated a large number of questions that come under the broad heading of the first of the two sorts of questions that define philosophy. In addition to being a very controversial project, the attempt to ground ethical values on scientific facts is plainly an enterprise that would further vindicate science as the setter of philosophy's agenda.

The nature of logical reasoning and its role in all the sciences also reflects the conception of philosophy as the study of questions that science can't answer. All of the sciences, and especially the quantitative ones, rely heavily on the reliability of logical reasoning and deductively valid arguments; but science also relies on **inductive arguments**—ones that move from finite bodies of data to general **theories**. Yet none of the sciences can address directly the question of why arguments of the first kind are always reliable, and why we should employ arguments of the second kind in spite of the fact that they are not always reliable. The only way the sciences could vindicate their methods is by using those very methods themselves! After all, they don't have any other methods. But any such "vindication" of the methods of science would beg the question, by assuming what it was meant to prove. Imagine accepting a promise to pay back a loan merely on the strength of a promise that one always keeps one's promises. Insofar as there are questions about the nature of scientific reasoning, these are ones the sciences themselves cannot answer.

What If There Are No Questions Left Over when Science Is Finished?

Our definition of philosophy does justice to the history of the sciences, and apparently to the division of labor between scientific and non-scientific inquiry about values and norms. And it makes sense of why logic, metaphysics, ethics, and epistemology should constitute one discipline despite their heterogeneity: they all address questions raised but as yet unanswered by science. But if we consider the definition again, there is one challenge it must face. Recall, as we have defined philosophy, that the first set of questions philosophy deals with are the questions that science—physical, biological, social, behavioral—cannot answer now and perhaps may never be able to answer.

But suppose one holds that in fact there are no questions that the sciences cannot now or cannot eventually answer. One might claim that any question that is forever unanswerable is really a pseudo-question, a bit of meaningless noise masquerading as a legitimate question, like the question "Do green ideas sleep furiously?" or "When it's noon GMT, what time is it on the Sun?" or "Did the universe and everything in it just double in size, charge, and every other physical magnitude?" or "How can we prove the universe and everything in it was not created five minutes ago?" Scientists and others impatient with the apparently endless pursuit of philosophical questions that seem to result in no settled answers may hold this view. They may grant that there are questions the sciences cannot yet answer, such as "What was happening before

the Big Bang that began the universe?" or "How did inorganic molecules give rise to life?" or "Is consciousness merely a brain-process?" But, they hold, given enough time and money, enough theoretical genius and experimentation, all these questions can be answered, as well as every other real question. The only questions science will leave unanswered, at the end of inquiry, will be pseudo-questions, ones that intellectually responsible persons need not concern themselves with. Of course, sapient creatures like us may not be around long enough in the history of the universe to complete science. But that is no reason to conclude that science and its methods cannot in principle answer all meaningful questions.

The claim that science can do so, however, needs an argument. The fact that there are questions like "What is a number?" or "What is time?" which have been with us unanswered for centuries is surely some evidence that serious questions may remain permanently unanswered by science. Could these really be pseudo-questions? We should only accept such a conclusion on the basis of an argument or a good reason. Suppose one wanted to argue that any question still left over at the "end of inquiry," when all the facts are in, must be a pseudo-question. As philosophers we can think of some arguments in favor of this conclusion. But these arguments that we can think of all have two related features: first they draw substantially on an understanding of the nature of science that science cannot itself provide; second these arguments are not ones science can construct by itself; they are philosophical arguments. And this is because they invoke normative premises, and not just the factual ones of science.

For example, the argument that questions science can never answer are really pseudo-questions (that science has no obligation to address) trades on the assumption that there are some questions science *should answer and does have an obligation to attend to*. But how do we decide what science *should* address? Presumably it should address those matters about which knowledge is at least possible. But then the responsibilities of science will turn on the nature, extent, and grounds of knowledge. Yet this is a matter for epistemology—the study of the nature, extent, and justification of knowledge. This means that philosophy is unavoidable, even in an argument that there are no questions science cannot answer, either now or eventually, or perhaps just "in principle."

Notice that this is not a conclusion that philosophers have some sort of special standing or perspective from which to ask and answer a range of questions that scientists cannot consider. These questions about science, its scope and limits, are as much questions that scientists can contribute to answering as they are questions for philosophers. Indeed, in many cases, as we shall see, either scientists are better placed to answer these questions, or the theories and findings they have uncovered play an essential role in answering them. But the important thing to see here is that philosophy is inescapable, even by those who hold that in the end all real questions, all questions worth answering, can only be answered by science. Only a philosophical argument can underwrite this claim. That argument, and the arguments to the contrary, require all the resources of the philosophy of science.

Furthermore, as much work in the philosophy of science suggests, it is by no means clear that there is a real distinction between philosophical questions and scientific ones, especially those raised at the moving frontiers of the sciences. Later in this book, we shall in fact explore some compelling arguments for this very conclusion. This means that on the definition we have advanced, we can expect important scientific contributions to perennially philosophical questions.

A Short History of Philosophy as the Philosophy of Science

Locating philosophy by its relationship to science relies as much on the history of philosophy as it does on the history of science.

At least since Descartes, early in the seventeenth century, the agenda of metaphysics and epistemology in European philosophy has been set by science, at first by physics and mathematics, later by the life sciences.

Descartes is famous for introducing his method of systematic doubt in order to fix the secure foundations of knowledge. He was equally interested in developing a physical theory that had no room for purposes, ends, goals, or any kind of "teleology" (from the Greek *telos*, meaning end toward which things strive), and could be expressed in mathematical equations. In this he was influenced by the great successes of Kepler and Galileo in expressing regularities about nature mathematically. Instead of statements about the purpose of things, he instead sought mechanistic laws, and formulated three such laws that bear a strong resemblance to those with which Newton later launched the revolution in physics. Descartes' rejection of purpose was of course not based on a successful non-teleological theory but on a philosophical argument that purposes did not exist in physical nature. Descartes had been convinced by "corpuscularianism," the view that the fundamental constituents of the universe are impenetrable "atoms" and that all physical processes are the result of the motion of these atoms and their collisions, fusions with, and fissions from one another. This "metaphysical" theory about the nature of reality goes back to Democritus among the Greeks and forward through the centuries to atomic theory. It is worth noting that Richard Feynman, arguably the most influential American physicist of the second part of the twentieth century, identified this notion that reality is atomic as the most important idea in physics. This made Descartes and all the corpuscularian philosophers of the seventeenth century prescient!

There were two great problems for this corpuscularian theory: First, gravity—a force that nothing can be shielded from, and that appears to be transmitted at infinite speeds through perfect vacuums. Gravity, thus, could not be carried through a vacuum by corpuscles or atoms since a vacuum is empty of everything, including atoms. Much of Descartes' work was directed at a solution to this problem. He never found one.

Second to gravity was the question of how we can have knowledge of these corpuscles or atoms given the fact that we cannot observe them. Much of John Locke's writings later in the seventeenth century were devoted to answering the

epistemological question of whether and how we can know anything about these atoms. Locke held that unobservable atoms and combinations of them caused our experiences and that some of the features of our experience resemble the properties of these atoms and the things made of them. The ones that Locke believed resemble the features of the objects—collections of corpuscles that cause our experiences—just happen to be the ones that seventeenth-century physics invoked: size, shape, matter. By contrast, Locke held that features of our experience like color, smell, taste, and texture were subjective properties of our experience that did not resemble the real properties of things that caused these experiences. Of course Locke did not appeal to physics in arguing for this claim. He thought he had deeper philosophical arguments for why our experiences represented reality at least in some of their features. Whence came the name for Locke's epistemology, "representative realism."

Few scientists and philosophers were satisfied with Locke's theory since there was no way to compare our experiences to their causes in order to determine whether the former resembled the latter. In fact, they held, Locke's solution to the problem of how we can have scientific knowledge about the world beyond our immediate observations simply encouraged skepticism about the claims of science. The general dissatisfaction with Locke's view sent philosophers and scientists in two different directions for their epistemologies, theories of the nature, extent, and justification of knowledge: **empiricism** and **rationalism**.

These two theories of knowledge will come up repeatedly in this book, so it is worth briefly explaining them and their differences. The original debate between these two epistemologies was about whether we had innate knowledge about the world and its nature. Descartes championed innateness. Locke repudiated it. The dispute about innateness seems to be a psychological one: is some of the knowledge everyone grants that people have somehow already inside their heads, known before birth, or is it all learned after birth by a mind that comes into the world as a blank slate, a *tabula rasa*? Following Descartes, rationalists in psychology and philosophy held the former view, empiricists the latter.

Over time, at least among philosophers, this disagreement moved from being a disagreement about the causes of our beliefs and knowledge to a dispute about the grounds or justification of knowledge. Empiricists held that all beliefs about the nature of things are justified by experience, observation, the collection of data, and the contrivance of experiments. Rationalists held that at least some beliefs, usually the most important, general, and fundamental ones in science, are justified by "pure reason" alone, and cannot be grounded on experience. Such beliefs are known to be true by the dint of the mind's employment of its own unaided powers of (presumably logical) thought. Notice the asymmetry here, which will continue to be important. Empiricists hold that anything we know about the world requires justification by experience. Rationalists by and large agree that most of what we know does require justification by experience. Theirs is not the contrary thesis that everything we know is justified without experience. Rather, they hold that in addition to a lot of knowledge that is justified by

experience, some knowledge is justified in other ways, and this knowledge is of particular importance to science.

It should be obvious from the role of experiment, observation, data collection, etc., that some type of empiricism is the "default" or "official" epistemology of science, as we shall call it throughout this book. Most rationalists will not demur from this label, provided we accept that there are at least a few claims of the profoundest significance in science that we know to be true but that cannot be justified by experience. Such statements, if there really are any, will have to be grounded in some other way, and will vindicate rationalism.

The key point for our purposes is to keep in mind that the dispute between rationalism and empiricism is a disagreement about the nature and extent of the kinds of grounds or justification that science enjoys.

Seeing that Locke's empiricism couldn't justify knowledge of unobservable corpuscles, but unwilling to adopt rationalism, Newton and many other scientists simply turned their backs on the philosophical problem of justifying corpuscularianism. This had the added convenience that, as Newton noticed, they would not have to solve the problem of reconciling the existence of gravity with the corpuscularian "philosophy." When asked to explain how gravity is possible, Newton famously said, "*Hypotheses non fingo*," Latin for "I make no hypotheses."

By the end of the seventeenth century, the philosopher George Berkeley was already attempting to articulate this view, treating physical theory not as knowledge of reality, but as a set of instruments or heuristic devices. Berkeley argued that we need not take the claims of physics about corpuscles or atoms seriously and decide on their truth or falsity. Rather we should treat theories as rules we devise for organizing our experiences and making predictions about what experiences we will have in the future.

The second of the two sorts of reactions to Locke's "representational realism" was to be found among the rationalist philosophers, especially Leibniz (early in the eighteenth century) and Kant (at the end of that century), who were eager to avoid skepticism about science at all costs. Each considered that Newton's laws of motion were so predictively powerful and the theoretical package they made up was so explanatorily compelling that they just had to be true, indeed necessarily true. But experience alone could never establish the **necessity**, still less the universality of these laws of nature. So **Newtonian mechanics** had to have a non-experimental grounding. Leibniz's grounding was roughly that when science was complete, it would turn out that there was only one logically possible set of laws of nature compatible with one another. This had to be the set of laws chosen by the omniscient deity when he made real this best of all possible worlds.

Kant had a quite different and far more influential strategy for establishing the certainty of Newton's laws. It's worth providing a few of its details since the way Kant set up the problem remained influential long after his epistemology fell out of favor.

Kant distinguished between what we know *a priori*, without experience, and what we can know *a posteriori*, only by having experience. Statements that are

necessarily true, like the propositions of mathematics, had to be *a priori*, since experience can never certify any statement as necessarily true. Statements such as "the number of planets is eight" are *a posteriori*—known only by experience and observation. Though Kant talked of *a priori* and *a posteriori* knowledge, this was not a distinction between innate versus learned beliefs, but a distinction between two types of justification of beliefs. We may acquire a belief like "Every event has a cause" through experience, but if it is known to be true, then the justification that turns it from mere belief to knowledge must be *a priori*, since no amount of finite experience could justify this statement about all events, past, present and future, observed and unobserved.

Kant also distinguished between statements that are true, roughly by definition, like "magnetic fields attract iron," and statements that are made true by facts about the world, like "magnetic fields are produced by the motion of electric charges." The former he called **analytic truths** and the latter, **synthetic truths**. Knowledge of the truth of analytic statements is no problem, since all we need to know is the meaning of the words involved to establish their truth—magnets are by definition iron attractors. Since no experience of the world is required to establish the truth of an analytic statement, it can be known *a priori*.

The problem for Kant was synthetic statements such as Newton's inverse square law of gravitational attraction:

$$F = G\,m_1 m_2\,/\,d^2.$$

This law asserts that the force of gravity falls off as the square of the distance between objects of mass m_1 and m_2. This statement is a synthetic truth; after all, the force of gravity could have varied as the cube of the distance or the square root of the distance, or any other function of the distance. It's a fact about our world that the force of gravity varies as the square of the distance. That makes the inverse square law a synthetic truth, not an analytic one. Now, this would be no problem, Kant thought, if Newton's laws were just our best guesses, based on experience, of what happens in the world, since guesses could be mistaken. But, Kant held, Newton's laws are universally and necessarily true. And experience could never ground universal or necessary truths. If Newton's laws are true, they had to be known—justified *a priori*, not justified by experience! How could this be possible? How could statements made true by facts about the world be known by us without any experience of those facts?! That was the great problem that Kant wrestled with in his famous work, *The Critique of Pure Reason*, written in 1781.

Kant's solution was that our minds are somehow structured in such a way that our thought imposes Newtonian concepts on our experiences. Independent of experience our minds are organized to think about the world in the framework provided by Newtonian mechanics. We necessarily impose the framework on the world—that's what makes it *a priori*. Since it's logically possible that we could have imposed a different conceptual scheme on the world, the Newtonian

framework we impose expresses a body of synthetic truths. So, Kant held, fundamental scientific knowledge is in fact *a priori* knowledge of synthetic truths. Kant's arguments for this revolutionary view were so difficult, so influential, and so interesting that scholars have been trying to understand them for more than 200 years. Meanwhile, advances in physics rather pulled the ground out from beneath Kant: The discovery of the special and general theories of relativity, along with quantum physics, showed that Newtonian mechanics was not necessarily true. It was not even true. So seeking to prove that it could be known *a priori* became pointless.

Only a few years before Kant wrote *The Critique of Pure Reason*, David Hume advanced a far different account of science, which held that none of its laws and theories are necessary truths, due to the fact that we can only justify them on the basis of experience. Hume was of course an empiricist, one who, by contrast to rationalists, was not nearly as eager to refute skepticism as rationalists were and are. He had no trouble with the notion that scientific knowledge can only be justified by experience, a conclusion he in fact claimed to draw from Newton's experimental methods.

Hume drew a distinction similar to Kant's between analytic and synthetic truths (though he did not use Kant's labels, which are still retained in contemporary philosophy). Unlike Kant, he argued that all *a priori* truths were analytic—true in virtue of the meanings we accord to the words that express them, and that all synthetic truths are *a posteriori*, known only through experience. Thus, while Hume agreed with Kant that all mathematical truths were necessary and so *a priori*, he held, against Kant, that they were all definitions or the logical consequences of definitions. Since experience can never establish the necessity of any statement, no scientific hypothesis about the world could be necessary. In fact, Hume held, we can never establish their truth finally and completely. Hume, by contrast to Kant, held that the fallibility of scientific hypotheses and theories was inseparable from their ability to explain what actually happened in the world. The only infallibly necessary truths, those of mathematics, Hume held, could be known for certain only because they made no claims about the world at all, but simply reflected the way we chose to define words. Hume ended his *Enquiry Concerning Human Understanding* with a stirring peroration, denying intelligibility to anything that was neither true by definition, like math, or justified by experience, like science:

> If we take in our hand any volume of divinity or school metaphysics, for instance; let us ask, Does it contain any abstract reasoning concerning quantity or number? No. Does it contain any experimental reasoning concerning matter of fact and existence? No. Commit it then to the flames: for it can contain nothing but sophistry and illusion.

For Kant and the rationalists the great difficulty was to show how any evidently synthetic truth about nature could be true *a priori*. For Hume and the empiricist philosophers who followed him, the problem was to show that all the

necessary truths of mathematics were really just disguised definitions and their consequence, and made no claims whatever about nature.

There are many truths of mathematics that seem to defy treatment as definitions. The classic example is the fifth axiom of Euclidean geometry. This postulate tells us that parallel lines never intersect and never increase their distance from one another. If this postulate were the consequence merely of the definitions of geometrical concepts, then its denial should have produced a contradiction or inconsistency somewhere in geometry. But no such contradiction has ever been derived from its denial. In fact nineteenth-century attempts to prove such contradictions merely led to the non-Euclidian geometries that were eventually employed by the theory of relativity that falsified Newton's laws and so refuted Kant. Even the recent proofs of several famous mathematical truths (Fermat's conjecture, the four-color theorem, Poincaré's conjecture) seem to rely on such complicated considerations that treating them as definitions or the consequences of definitions is implausible. This is important since *a priori* truths in mathematics, which appear to be synthetic statements about reality, greatly strengthened Kant's rationalist claim that the laws of physics and perhaps other parts of science are also synthetic *a priori* truths!

No matter who was right, rationalists or empiricists, the way philosophy was developing at the end of the eighteenth century, its preoccupation and symbiotic relationship with science was undeniable.

For a variety of reasons, some of them having to do with the end of the Enlightenment, the excesses of the French Revolution and the advent of Romanticism in the nineteenth century, European philosophy lost the interest in physics and mathematics that had been its stock in trade for 250 years or so. The agendas of philosophers like Hegel and those who followed him were not driven by an interest in science, its view of reality and the methods by which it advanced. In fact, it is fair to say that Hegel and his nineteenth-century European followers sought to substitute a speculative philosophy of nature for the claims made by Newton and his successors. Unlike the tradition that produced Kant, their aim was not to provide a philosophical foundation for what science had achieved. From the early 1800s on, Hegel sought to describe a "reality" that could never be grasped by mere experiment or mathematically expressed theorizing. The resulting century-long tradition of speculation uncontrolled by experiment and mathematics produced a large number of books with ever-diminishing influence among scientists and ever-increasing perplexity among those mainly English-speaking philosophers who still put any confidence in science to reveal the nature of reality. An example of Hegel's writings on physics makes the problem of comprehension clear:

> Matter in its first elementary state is pure identity, not inwardly, but as existing, that is, the relation to itself determined as independent in contrast to the other determinations of totality. This existing self of matter is light.
>
> As the abstract self of matter, light is absolutely lightweight, and as matter, infinite, but as material ideality it is inseparable and simple being outside of itself.

In the Oriental intuition of the substantial unity of the spiritual and the natural, the pure selfhood of consciousness, thought identical with itself as the abstraction of the true and the good is one with light. When the conception which has been called realistic denies that ideality is present in nature, it need only be referred to light, to that pure manifestation which is nothing but manifestation.

Heavy matter is divisible into masses, since it is concrete identity and quantity; but in the highly abstract ideality of light there is no such distinction; a limitation of light in its infinite expansion does not suspend its absolute connection. The conception of discrete, simple, rays of light, and of particles and bundles of them which are supposed to constitute light in its limited expansion, belongs among the rest of the conceptual barbarism which has, particularly since Newton, become dominant in physics. The indivisibility of light in its infinite expansion, a reality outside of itself that remains self-identical, can least of all be treated as incomprehensible by the understanding, for its own principle is rather this abstract identity.

(Hegel, *Philosophy of Nature*, "Elementary physics," sections 219, 220)

It is probably evident how scientists even in the nineteenth century reacted to this sort of philosophy. Towards the end of the nineteenth century philosophers in Europe and the English-speaking countries began to repudiate speculation uncontrolled by science as an acceptable philosophical method.

There were two catalysts for this sea change that turned much philosophy back to the tradition of taking science seriously, which had dominated the discipline from Descartes to Kant. One was an achievement in logic. The second was a philosophically inspired scientific breakthrough.

From the time of Aristotle, philosophers, mathematicians, scientists, and for that matter lawyers, had recognized many different "syllogisms"—forms of deductively valid proof such as: all *a* are *b*, all *b* are *c*, therefore all *a* are *c*. There are exactly 256 different syllogistic forms of argument. These arguments are all "truth-preserving." If their premises are true then their conclusions will be true. Their importance consists in this truth-preservingness. It means that they cannot lead us astray in reasoning: if one starts with true premises and correctly uses one of these forms of argument, one's conclusions are guaranteed to be true as well.

The trouble with logic was the complete absence of a theory that explained why these forms of argument are valid—what feature they share in common that confers validity on their instances. Around 1900, independently, the British philosophers Bertrand Russell and Alfred North Whitehead, along with the German philosopher Gottlob Frege, developed a system of symbolic logic that solved the problem: it not only explained the validity of the 24 valid forms of syllogism known since antiquity, but also provided the basis for establishing the validity of much more complicated forms of deductive reasoning used throughout mathematics and science.

This discovery was an achievement on a par with what atomic theory did for the periodic table. Atomic theory established the rightness of the periodic table of the elements by explaining the table from facts about electrons, protons, and neutrons. Similarly, Russell, Whitehead, and Frege's discovery explained the correctness of all deductively valid forms of argument—including the 256 syllogisms—by deriving all of its forms from a small number of basic axioms. Even better, these axioms appeared to be trivially true (truth-functional) definitions of the logical constants—"and," "or," "not," "some," and "all," or the immediate consequences of these definitions. Russell and Whitehead exploited this feature of their new logical system to try to derive set theory, number theory, mathematical analysis and other parts of mathematics from the definitional beginnings of logic alone. This last endeavor failed, but did not detract from their original achievement.

Looking back at the work of Hume from the vantage point of this discovery about logic gives great new strength to the empiricist program in philosophy. Suddenly, a solution to the great problem facing the empiricists—the problem of showing all mathematics to be analytically true—seemed to be within reach. Recall, unless empiricists could show that the *a priori* truths of mathematics were analytic, these truths provided an example of *a priori* knowledge that could be synthetic. This allowed rationalists to hold out hope that they could show that physics and perhaps much more of science was also necessarily true despite its synthetic character, despite the fact that it describes facts about reality.

The second important event for twentieth-century philosophy was Albert Einstein's discovery of the special and general theories of relativity, and its displacement of Newtonian mechanics as the fundamental theory in physics. To begin with, Einstein himself identified the intellectual sources of his discoveries not in the experimental work of physicists but in the analysis and critique of Newton's theories of space and time by two eighteenth-century philosophers, the rationalist Leibniz and the empiricist Berkeley. The lesson he drew from their work was that even the most fundamental concepts in science—length, time, velocity, mass—needed to be defined by experiences, empirical manipulations and operations, and not in terms of the abstract, ambiguous, and ornate verbiage of Hegel's definition of matter or light illustrated above. In short, Einstein showed that the foundations of physics require the empiricist epistemology Hume had advanced in his account of science.

Thus Hume's attack on "school metaphysics"—philosophy unconnected to science—found a new life in the first half of the twentieth century. It was combined with the discoveries of Russell, Whitehead, and Frege into a philosophy called "**logical positivism**" or "**logical empiricism**." Unsurprisingly, given Einstein's and Frege's influence in the German-speaking countries, logical positivism began life in post-World War I central Europe as the philosophy of a group called "the Vienna Circle."

The Vienna Circle would have been very sympathetic to the definition of philosophy in terms of questions that the sciences cannot answer and questions about why the sciences cannot answer them. But with regard to questions

that the sciences cannot answer, it would have denied that there were any such questions. The very idea of "pseudo-questions" in fact arose among the logical positivists. Because logical positivism denied that meaningful questions would remain when science is complete, it of course did not concern itself directly with the second category of questions. Logical positivism had an argument for such disparate treatment.

The argument that all meaningful questions would ultimately be answered by scientific (empirical) inquiry rested on two distinct theses. One of these was a so-called "Principle of **verification**"—the claim that to be scientifically intelligible, or as they said, "empirically meaningful" or "cognitively significant," a set of concepts or words had to make a specific testable claim about experience, something that could be subject to empirical observation or experimental test. This immediately ruled out as scientifically meaningless most of the Hegelian philosophy of the nineteenth century, all of theology, aesthetics, and a large number of other areas of apparent inquiry no one supposed the sciences could address. It also ruled out ethics, political philosophy, and all other "normative" statements as meaningless too (no surprise if science cannot answer any question about normative matters). The only questions left as meaningful were the ones empirical inquiry—i.e. science—could address. Thus there were no questions for philosophy or any other non-science discipline.

If the positivists had been able to formulate a precise principle of verification that could have served as a litmus test for statements that were meaningful and those that were not, it could have been used to distinguish real science from what the positivists and their philosophical (not political) "fellow travelers" (Karl Popper for instance—for more on Popper see Chapter 11) called "pseudoscience." Demarcating real science from discourse that merely pretended to be scientific in order to secure the prestige and influence science carries became a powerful motivation for some philosophers attempting to frame a principle of verification. Unfortunately for their program, they were never able to do so (for reasons discussed in detail in Chapter 8). The difficulty they faced was twofold: First, their own self-imposed requirement that the principle be unambiguous, precise, and give a definite answer about each statement to be assessed for empirical meaningfulness; second that it give the "right answer" for the statements that physicists accepted as empirically meaningful, including statements about completely unobservable things and properties like electrons, protons, their charges, masses and constituents, if any. The logical positivists were never able to satisfy themselves that any of their many increasingly complicated candidates for a principle of verification could satisfy these two requirements. This problem, and their unswerving intellectual honesty about it, sowed the seeds of logical positivism's downfall. In the next chapter we'll consider why the "demarcation problem" remains important even if we cannot find a litmus test for science vs. non-science.

The other of logical positivism's tenets was that the claims of logic and mathematics were not testable but not meaningless either. They were, to use Kant's

label (which remains in use), analytical truths—definitions and the consequences of definitions. Here the positivists had the assurance of Russell, Whitehead, and Frege's great breakthrough in symbolic logic, and its promise eventually to show in detail how math was just the consequence of definitions. The logical positivists held that philosophy, as they practiced it, was to be classed together with logic and mathematics as analytical propositions—definitions and their consequences. Like the remarkable results of mathematical research, philosophy only looks novel to us because we are not logically omniscient. The true and correct philosophy, like surprising new theorems in math that no one ever thought of before, was always there implicit in the premises with which inquiry began. It was just waiting for someone smart enough to draw it out and show us that we were already committed to it.

Looking back at the history of philosophy, the logical positivists were not surprised to be able to find that much of the history of philosophy from Descartes to Kant could be interpreted as attempts to offer such analytically true claims—mainly the epistemology of science and mathematics, and analysis of the meaning of the terms in which scientific theories are expressed. This enabled them to treat philosophy as an intellectual enterprise whose ultimate aim was to show that science can answer all empirically meaningful questions, and that philosophy's only role was, like that of mathematics, to provide clear definitions and logical rules that would enable everyone to see why only science has "cognitive significance."

The Vienna Circle of logical positivists and other philosophers sympathetic to them were active defenders of democracy, advocates of social justice, and opponents of totalitarianism. Therefore they had to flee European fascism in the late 1930s. They recognized that Soviet communism was equally threatening to their philosophy and their lives. As a result, by the 1940s logical positivism had moved mainly to the United States. There it flourished for a long time, and eventually vindicated the traditional association between philosophy and science that is reflected in our original definition. But in its development of the philosophy of science as a sub-discipline of its own, logical positivism sowed the seeds of its own displacement in the second half of the twentieth century. We will return to some of these themes at the beginning of Chapter 2, and describe how logical positivism set the agenda of the philosophy of science that remains in place, even after its eclipse.

Meanwhile, the take-home lessons of this history are twofold: first we must demand of any alternative definition of philosophy that tries to loosen its connection to science, that it do at least as good a job as our definition in making sense of the history of science and the history of philosophy over the last 400 years. Second, we must recognize that deciding what the scope and limits are of science's ability to answer all our questions is no easy matter, and that the difficulty of doing so makes philosophy an indispensable inquiry.

Summary

Philosophy and science are inescapable for anyone who is interested in either subject. In fact, the agenda of philosophy throughout its history and at present cannot be understood except as a reflection of problems left to it by the sciences as they established their independence of it. And the history of science, especially the history of its successes since the sixteenth century, is in significant part a matter of acquiring confidence in a particular metaphysics and epistemology.

The major figures of European philosophy disputed whether empiricism or rationalism could account for the success of science since Newton. These two different theories of knowledge continue to contend with one another, even though empiricism has come to be more widely accepted, among scientists at any rate. Philosophers would like to accept it, but, as we will see throughout this book, they recognize problems about the nature of science and its methods that empiricism cannot seem to cope with. These are the problems that make the philosophy of science so central to philosophy as a whole.

Study Questions

Answering the study questions at the end of each chapter does not simply require a recapitulation of information provided in the chapter. Rather, these raise fundamental questions about philosophical theories aired in the chapter, and identify controversial issues on which readers are invited to disagree with the authors, bring up examples, arguments and other considerations on which the text is silent, and make up their own minds. Some of the questions raised at the end of each chapter are worth revisiting after reading subsequent chapters.

1. This chapter offers a potentially controversial definition of philosophy. Provide an alternative definition for philosophy, which accounts for the unity of the disparate parts of the discipline: metaphysics, epistemology, logic, ethics and political philosophy, aesthetics, etc.
2. Construct an argument to show that any question that science cannot in principle ever answer, even when all the facts are in and science is complete, is a pseudo-question.
3. Defend or criticize: "The empiricism vs. rationalism debate may advertise itself as one about justifications but it still comes down to the original dispute between Descartes and Locke about whether there are innate ideas."
4. Hume condemned "school metaphysics and divinity" as "sophistry and illusion." What contemporary substitutes would he have the same intellectual contempt for these days?

5. "Corpuscularianism was mere philosophy. Atomism is science." Discuss.
6. Are there philosophical questions whose answers have no relevance for philosophy of science, or no relevance for science? Why?

Note

1 Words printed in the text in bold type are defined in the Glossary.

Suggested Readings

Readers seeking an introduction to the history of science, and especially its history since the Renaissance, will profit from Herbert Butterfield, *The Origins of Modern Science*. Thomas Kuhn, *The Copernican Revolution*, provides an account of seventeenth-century science by the historian of science most influential in its philosophy. I. Bernard Cohen, *The Birth of a New Physics*, and Richard Westfall, *The Construction of Modern Science*, provide accounts of Newtonian mechanics and its emergence. James B. Conant, *Harvard Case Histories in the Experimental Sciences*, is another influential source for understanding the history of the physical sciences.

Hans Reichenbach, one of the most important twentieth-century philosophers of science, traces the influence of science on philosophy in *The Rise of Scientific Philosophy*. A classic work in the history of scientific and philosophical ideas is E. A. Burtt, *Metaphysical Foundations of Modern Science*, first published in 1926.

An excellent and reliable contemporary introduction is Steven Shapin, *The Scientific Revolution*. *The Cambridge History of Science*, especially volume 3, *Early Modern Science*, edited by Katharine Park and Lorraine Daston, and volume 4, *The Eighteenth Century*, edited by Roy Porter, provides a great deal of the best recent scholarship on the history of science and its philosophy in its most formative period.

J. L. Heilbron, *The History of Physics: A Very Short Introduction*, gives a rapid tour through the history of this discipline.

2   Why Is Philosophy of Science Important?

Overview

There are urgent practical reasons to try to sort out any differences between scientific knowledge and other sorts of knowledge. Matters of public policy and individual well-being depend on being able to tell science from pseudoscience. However, doing so is not as easy as many might hope.

Besides calling upon philosophy to help distinguish it from other human undertakings, science continues to have a profound impact on the agenda of philosophy. Its impact on the rest of culture is equally significant, and its impact across cultures seems uniquely universal. If this is true, it requires explanation.

Scientific Questions and Questions about Science

Among the questions science cannot yet answer are questions about why the sciences cannot (or perhaps will never be able to) answer all questions. Call the questions about what a number is, or what time is, or what justice is, first-order questions. Second-order questions, about why science cannot as yet cope with the first-order questions, are themselves questions about what the limits of science are, how it works, how it is supposed to work, what its methods are, where they are applicable, and where they are not. Answering these questions will

either enable us to begin to make progress on the hitherto unanswered first-order questions, or enable us to recognize that some of these first-order questions are not ones that science can or needs to answer. Answering questions about what the nature of science is and what its methods are can also help us assess the adequacy of proposed answers to scientific questions.

But there are other concerns—not directly scientific ones—where the philosophy of science may be able to help us. Here are some important examples:

Philosophers, scientists, and other defenders of the integrity of science and its uniqueness as an instrument for the acquisition of objective knowledge have long opposed granting equivalent standing to non-scientific ways of belief formation. They have sought to stigmatize astrology, dianetics (and scientology), tarot cards, telekinesis, "creation science" or its various latter-day variants, "intelligent design theory," or for that matter any New Age fashion, Eastern mysticism, holistic metaphysics, as pseudoscience, distractions, diversions, and unworthy substitutes for real scientific explanation and its application in practical amelioration of human life. As we saw in the last chapter, logical positivists hoped to be able to tell science from pseudoscience by use of a simple litmus test. For reasons we will explore in Chapter 8, they failed to create an easy rule that could do this. So the problem remains. What is needed is an account of exactly what makes for scientific knowledge, and a decision about whether there are other kinds of knowledge besides what science provides.

Human credulity, the search for and belief in satisfying narratives, conspiracy theories, simple solutions and miracle cures, and shortcuts to knowledge and wealth, is so strong that it continues to be exploited even 350 years after the scientific revolution. Indeed, in some respects the situation is worse than it was a hundred years ago.

In the United States, an alliance was formed in the 1990s among groups of people impatient with the slow progress of orthodox empirical, controlled double-blinded experimental laboratory-based science to understand and deal with illness, together with those convinced that there was important, therapeutically useful knowledge about illness, its causes and cures, embedded in one or another non-experimental approach. This alliance prevailed upon the U.S. Congress to direct the experimentally-oriented National Institutes of Health to establish an Office of Alternative Medicine mandated to spend significant sums of money (allegedly diverted from the funding of mainstream orthodox scientific research) in the search for such knowledge. Members of this alliance often argued that there are some therapeutic substances that only work when employed under the condition that the patient and/or physician know that the patient is treated with these drugs and furthermore believe in their effectiveness. On their view a controlled experiment, in which neither patients nor physicians know whether the patient receives the drug or a placebo, cannot therefore be employed to test the efficacy of the treatment. If such a controlled double-blind experiment is the only way we can scientifically assess effectiveness, it follows that these claims about "alternative medicines" are beyond the reach of any scientific assessment. Hence, their advocates argue, the search of knowledge about such medicines cannot be scientific.

In the United Kingdom, during the first decade of the twenty-first century, a debate arose about the continued payment of increasingly scarce funds by the National Health Service for homeopathic treatments. (Homeopathy is the theory that illnesses can be cured by the administration of minute quantities of substances that cause symptoms similar to the illness being treated.) Even after controlled experiments showed that their effectiveness was no better than that secured by (much cheaper) placebos, and after analytical chemists demonstrated that the medicines advocated by homeopaths are undetectable in the dilutions required by the theory, its advocates—especially those practitioners whose services were paid by the National Health Service—continued to insist that their theory and treatments were scientific. Meanwhile, physicians and other scientists denied this claim and were subject to libel lawsuits in the British courts.

It is obviously difficult for opponents of this diversion of scarce resources from science to argue that alternative medicine cannot provide scientific knowledge, unless they have an account of what makes scientific findings unique.

On the other hand, advocates of such approaches have an equal interest in showing that it is in the nature of the orthodox scientific method to be blind to non-experimental knowledge. Such advocates can make common cause with others—humanists for example, who oppose what they call "scientism," the unwarranted overconfidence in the established methods of science to deal with all questions, and the tendency to displace other "ways of knowing" even in domains where conventional scientific approaches are inappropriate, unavailing, or destructive of other goals, values, and insights.

Both parties to this dispute have an equal interest in understanding the nature of science, both its substantive content and the methods by which it proceeds in the collection of evidence, the provision of explanations, and the appraisal of theories. In other words, both sides of the debate need a demarcation criterion they can agree on in order to settle their dispute.

Those who appreciate the power and successes of the natural sciences, and who wish to apply methods successful in these disciplines to the social and behavioral sciences, have a special incentive to analyze the methods that have enabled natural science to attain its successes. Since the emergence of the social and behavioral sciences as self-consciously "scientific" enterprises, these scientists, and some philosophers of science, have held that the relative lack of success of these disciplines, by contrast to the natural sciences, is due to a failure to correctly identify or implement the empirical methods that have succeeded in natural science. For these students of social science, the philosophy of science has an obviously *prescriptive* role. Once it reveals the features of evidence gathering, the explanatory strategies, and the ways in which both are applied in the natural sciences, the key to similar advances in the social and behavioral sciences becomes available. All the social and behavioral sciences need to do is employ the right method. Or so these students of scientific methodology argue.

However, there are opponents of the scientific treatment of social and behavioral issues. They wish to argue that the methods of natural science are inapplicable to their subjects, that "scientistic imperialism" is both intellectually

unwarranted and likely to do harm by dehumanizing personal relationships and fragile social institutions. They go on to hold that such an approach is likely to be misapplied to underwrite morally dangerous policies and programs (for example, various eugenic policies pursued by many countries during the twentieth century), or even to motivate inquiry into areas best left unexamined (such as the genetic basis of violence, criminality, mental illness, intelligence, etc.). It is clear that these defenders of the insulation of human affairs from the simple-minded and inapplicable empirical methods of the natural sciences need to understand what scientific inquiry consists in. These defenders of the epistemological integrity of the non-empirical approach to understanding human affairs need a theory of knowledge that shows how reliable knowledge can be acquired without using the methods of the natural sciences. This means that both they and those with whom they argue about the right way to proceed in the social sciences and social studies need to understand science, its nature, and its limits, if any.

The philosophy of science is thus an unavoidable inquiry.

Modern Science Has Implications for Philosophy

Besides the traditional questions that each of the sciences left as an intellectual legacy to philosophy, the development of the sciences over two millennia or more has persistently raised new questions with which philosophers have struggled. Moreover, these two millennia of scientific development have shaped and changed the agenda of philosophical inquiry as well. In Chapter 1 we saw that science was the most powerful source of philosophical inspiration due to its revolutionary successes in the seventeenth century through the nineteenth, and into the twentieth century. It remains so.

Consider the problem of free will. Newton showed that motion—whether of planets and comets, or cannon balls and tides—was governed by a small number of simple, mathematically expressible, and perfectly exceptionless laws. These laws were deterministic: given the position of the planets at any one time, the physicist could calculate their position at any past or future time. If Newton was right, a body's position and momentum at any one time fixes position and momentum for all times. What is more, the same inexorable laws bind all matter, anything with mass. The determinism of Newtonian mechanics raises the specter of determinism in human behavior as well. For if humans are nothing but complex collections of molecules, i.e. of matter, and if these collections behave in accordance with these same laws, then there is no real freedom of choice, but only the illusion of it. Suppose we trace the causes of our apparently free actions (for which we are responsible), back through their previous causes to our choices, our desires, and the physical states of our brains in which these desires are represented. If the brain is nothing but a complex physical object whose states are as much governed by physical laws as any other physical object, then what goes on in our heads is as fixed and determined by prior events as what goes on when one domino tile topples another in a long row of them. If the causes that fixed the events in our brain include events over which we have no

control—say, our upbringing, our present sensory stimulation and physiological states, our environment, our heredity—then it may be claimed that there is no scope in this vast causal network for real free choice or action (as opposed to mere behavior), and so no room for moral responsibility. What is determined by the prior state of things, and therefore beyond our control, is not something for which we can be blamed, or praised for that matter.

With the success of Newton's theory, determinism became a live philosophical option for human affairs. But it remained open to some philosophers, and of course to many theologians, to hold that physics does not bind human action, or for that matter the behavior of any living thing. They held that the realm of the biological was beyond the reach of Newtonian determinism. And the proof of this was that physical science could not explain biological processes at all, let alone with the power and precision with which it explained the behavior of mere matter in motion.

Until the middle of the nineteenth century, opponents of determinism might have comforted themselves with the thought that human action, and the behavior of living things generally, was exempt from the writ of such Newtonian laws of motion. Human action, and biological processes, are evidently goal-directed. Humans act for purposes that reflect the existence of pedestrian ends. Nature reflects the cosmic purposes and vast scheme of things that God affords. The biological realm seems to show too much complexity, diversity, and adaptation to be the product of mere matter in motion; its appearance of design shows the hand of God. Indeed, before Darwin, the diversity, complexity, and adaptation of the biological realm was the best theological argument for God's existence and for the existence of a "plan" that gives the universe meaning. This plan (of God's) was also at the same time the best scientific explanation for these three features of the biological realm.

It was Darwin's achievement, as the theologians who opposed him so quickly realized and so strenuously denounced, to destroy the grounds of this theologically inspired metaphysical world view along with its scientific explanation of adaptation in the biological realm. As Darwin wrote in his unpublished notebooks 20 years before he dared publish *On the Origin of Species*, "Origins of Man now proved. Metaphysics must flourish. He who understands baboon would do more towards metaphysics than Locke." We cannot summarize Darwin's alternative to revealed religion here (the matter is taken up again in Chapter 6 and at greater length in Chapter 9), but if Darwin's mechanistic, purpose-free account of diversity, complexity, and adaptation as the result of heritable genetic variation and natural environmental selection is right, there is a strong argument that nothing in the universe has any meaning, purpose, or intelligibility beyond the sort of clockwork determinism that Newton's discoveries revealed. And this is a profoundly philosophical conclusion, that goes beyond even determinism by showing that all purpose in nature is illusory. Between them Newton and Darwin were the great sources of philosophical materialism or physicalism, which undermines so much traditional philosophical theory in metaphysics, the philosophy of mind, and may still threaten moral philosophy.

The philosophical problems that both theories raised were still at the center of the philosophical agenda in the twentieth century.

But then, twentieth-century developments in physics and the foundations of mathematics made matters more complex and shook the confidence of philosophical materialism far more than any merely philosophical arguments. First, the attempt to extend deterministic physical theory from observable phenomena to unobservable processes came up against the appearance of sub-atomic indeterminism in nature. It has turned out that at the level of quantum processes—the behavior of electrons, protons, neutrons, the photons of which light is composed, electromagnetic radiation—any laws seem to be ineliminably indeterministic. It is not just that we cannot know what is going on with certainty and have to satisfy ourselves with mere probability. Rather, almost all physicists believe it to have been physically established that the probabilities of quantum mechanics explain the behavior of the fundamental constituents of matter (and so of everything), with the fantastic precision that they reflect. What is more, the way that these probabilities fall out in experiments precludes the existence of a deeper deterministic theory that somehow explains these probabilities.

We cannot explain here why physicists hold that these probabilities can never be eliminated, but we can illustrate them. Whether a single particular uranium atom will emit an alpha particle in the next minute has a probability of, say, 0.5×109. No amount of further inquiry will raise or lower that probability; there is no difference in the state of a uranium atom that results in alpha emission during one minute and in the state of the atom when it does not emit the particle during the course of another minute. At the fundamental level of nature, the so-called "principle of sufficient reason"—that every event has a cause—is continually violated.

Of course by the time electrons, protons, and other particles are put together into molecules, their behavior begins asymptotically to approach the determinism that Newtonian mechanics demands. But Newton turns out to have been wrong. In case one might hold out hope that the world of observable objects that Newton's theory deals with is exempt from quantum mechanical indeterminism, just recall that Geiger counters are observable detection devices whose clicking noises when held over radioactive materials are the result of quantum undetermined emissions of alpha particles making an observably detectable difference in the macro-world.

Now, does all this mean that if determinism is false, free will and moral responsibility are after all vindicated as acceptable components of our philosophical world-view? Things are not that simple. For if the fundamental sub-atomic interactions that constitute our brain processes are not determined by anything at all, as quantum physics tells us, then there is even less room for moral responsibility in our actions. If indeterminism is true, our actions will stem from events that have no causes themselves, no reason at all for their occurrence. In short, quantum indeterminacy deepens the mystery of how human agency, deliberation, real choice, free will and ultimately moral responsibility is possible. Suppose that we can trace your actions, both the morally permissible and

impermissible ones, back to an event, say, in your brain, which itself had no cause, but was completely random, undetermined, and inexplicable, an event over which neither you nor anyone else, nor for that matter anything else, had any control whatsoever. Well, in that case, no one can be morally responsible for the effects of that event, or its effects in and on your desires, your choices, your actions.

If the direction in which science carries philosophy is a one-way street towards physicalism, determinism, atheism, and perhaps even nihilism, then the intellectual obligation to science of those who wrestle with philosophical questions would be unavoidable. We must understand the substantive claims of physical science, we must be well informed enough to interpret the significance of these claims for philosophical questions, and we must understand the strengths and limitations of science as a source for answers to these questions.

But in fact, the direction in which science seems to carry philosophy is by no means a one-way street towards physicalism, determinism, atheism, and nihilism. Since the sixteenth century many philosophers and scientists have endorsed the argument of the mathematician, physicist, and philosopher René Descartes, that the mind is distinct from the body or any part of the body, in particular the brain. Descartes' followers have never argued that the mind can exist without the brain, any more than human life can exist without oxygen-respiration. But they have held that (just as life is not just the respiration of oxygen) thought is not identical to any brain process. The mind is a separate and distinct substance, a non-physical one, and therefore not subject to any laws that physical science can uncover. If the mind is not a physical thing, this may exempt humans and human action from the natural laws that science uncovers or even from scientific study itself. It may turn out that humans and human actions must be understood by methods completely different from those that characterize natural science. Or it may be that human affairs cannot be understood at all.

This view that the mind is non-physical and beyond the reach of natural science may be greeted with dismay and stigmatized as obscurantist, as an obstacle to intellectual progress. But calling it names will not refute the arguments that Descartes and others advanced on its behalf. And the general weakness of those social sciences inspired by methods and theories of natural sciences should give some further pause to those who reject Descartes' arguments. Can it really be that the only obstacle in social science to the sort of predictive precision and explanatory power we have in natural science is the greater complexity of human behavior and its mental causes?

Among those who answer this question in the affirmative have been psychologists and others who have sought to understand the mind as a physical device along the lines of the computer. After all, the neural architecture of the brain is in some important respects like that of a computer: it operates through electrical signals that switch nodes of a network to states of "on" or "off." Psychologists interested in understanding human cognition have sought to model it on computers of varying types, recognizing that the human brain is vastly more powerful than the most powerful super-computer and uses computational

programs quite different from those with which we program current computers. But, if the brain is a powerful computer and the mind is the brain, then at least modeling cognition by developing simple programs that simulate aspects of it on computers less powerful than the brain will show us something about how the mind works.

It is at this point that some argue that the development of science itself raises obstacles to this "scientistically" inspired research program. What we know for sure about computers is that they operate by realizing software programs with certain mathematical features. In particular, the software makes a computer operate in accordance with a finite set of mathematically expressed axioms. As a simple example, consider the arithmetical calculations that a computer is expected to make. It can multiply any two numbers. It can do any one of an indefinitely large number of calculations not because it is programmed with the correct answer to every multiplication problem, but because it is programmed with the rules of multiplication in the form of an axiom of arithmetic. Of course, there are limitations on the calculations that a computer can actually carry out. Anyone who has played with a calculator knows what some of them are. If it runs out of power, or if the numbers to be multiplied have too many places for the read-out screen, or if an illegal operation like dividing by zero is attempted, or if the machine is ordered to calculate pi, then it will not give a unique complete right answer. In this respect computers are like human calculators: they have breakdowns and failures.

But in the 1930s an Austrian mathematician, Kurt Gödel, proved mathematically that in a critically important way computers are not like human calculators. And subsequently some philosophers and scientists have argued that this result is an obstacle to a scientific understanding of cognition and of the mind. What Gödel proved was this: Any **axiomatic system** powerful enough to contain all the rules of arithmetic is not strong enough to provide its own completeness: that is, it is not strong enough to prove that every truth of arithmetic follows from its axioms. To provide such a system's completeness requires that we employ a stronger system, one with more or different axioms. And similarly for this stronger system, proving its completeness is beyond its reach as well. What is more, proofs of consistency will always be relative to some one or more stronger system in which the completeness of the weaker system can be provided.

Gödel's proof has two implications. First, it probably did more than anything else to undermine the logical positivists' confidence in their own account of science as purely empirical and experimental. The reason is subtle. Recall from Chapter 1 that, like most empiricists since Hume, the logical positivists had solved their problem of how, in the case of mathematics, we can have *a priori* knowledge, and knowledge of necessary truths at that. They were able to "explain away" our *a priori* knowledge of these necessary truths, without admitting that science conveyed *a priori* truths as well, by arguing that all mathematical truths were "merely" definitions and the consequences of definitions. So, mathematics was necessarily true but didn't convey new knowledge of facts about reality. All its truths were *a priori* truths because they were analytic truths.

Following Russell, Whitehead, and Frege, the logical positivists hoped to show this by deriving them from the axioms of logic, which were themselves analytic truths.

Gödel's proof showed that this ambition was unattainable. By showing that no axiomatic system strong enough to derive even the truths of arithmetic could be both consistent and complete, he demonstrated that the logical positivist account of mathematics as analytic truths had to be mistaken. Thus, like empiricists before their time, the logical positivists had either to find a new account of how, in mathematics, we can have knowledge of necessary truths or else give up their empiricist epistemology. The increasing realization of this problem was one of the most powerful reasons that philosophers of science began to surrender logical positivism.

The second of the two implications of Gödel's proof is potentially of much wider significance. It provided the basis for an argument against the notion that human cognition is in any interesting sense a matter of computation, or that the mind works in any way like a computer. From this conclusion some philosophers and even some scientists have been tempted to move to full-blown dualism, denying that the mind is the brain at all.

It is obvious that the human mind embodies an understanding of arithmetic that is not limited in the way that a computer is. After all it was a human mind that proved that computer programs for mathematics could not be complete and consistent. Unlike a computer, the human mind's "representation" of arithmetic is not axiomatic. If so, it is not a material thing, since no material thing, no machine, can carry all mathematical knowledge. Whether the human mind grasps arithmetic axiomatically or not, there is a further aspect of Gödel's proof to consider. If an axiomatic system is provably consistent—i.e. contains no contradictions, no necessary falsehoods—then, Gödel showed, there will always be at least one expression formulable in the language of the consistent system that is undecidable in that system. That is, the consistent system is incomplete. Gödel's strategy was roughly to show that for any consistent system at least as powerful as arithmetic, there is always a true sentence of the form "this sentence is not provable in the system" that is unprovable in the system. No axiomatic system of the sort programmed on any computer capable of arithmetic can be both provably complete and consistent. Since the last thing we want is a computer or a calculator that is inconsistent—that generates false answers to calculations—we must reconcile ourselves to computers whose programs are not provably complete.

But, apparently, this is not a limitation on us. To begin with, we humans, or at least one of us, proved this result. Kurt Gödel was able to do so because, unlike computers, minds like ours can identify the inconsistent statement in one axiom system-program that is complete, and minds like ours can identify the one true statement that is unprovable in the closest alternative axiom system-program that is consistent. So, evidently our minds, or at least the rules of thought that we employ, are not merely the software implemented on the hardware (or wet-ware) of our brains. Since this mathematical result reflects a limitation on any physical system, no matter what material it is made from—silicon chips,

vacuum tubes, cogs and wheels, or neurons and synapses—it is argued, by some distinguished physicists among others, that the human mind cannot be material at all. And therefore, it is not subject to study by means appropriate to the study of material objects, whether those means are to be found in physics, chemistry, or neurobiology.

Here then is a result of modern science (and mathematics) that tends to undercut the confidence of the purely scientific world view as a philosophy. Readers should be warned that the conclusions drawn above from Gödel's "incompleteness" proof, as it has come to be known, are highly controversial and by no means widely shared. Indeed, we do not accept the proof as showing anything like the conclusions drawn above. But the point is that results in science like this one are of overwhelming importance to the traditional agenda of philosophy, even when as in this case they suggest limitations on the scientific worldview as a philosophy.

The Cultural Significance of Science

Whether we like it or not, science seems to be the only universally welcome contribution of European civilization to all the rest of the world. It is arguably the only thing developed in Europe that every other society, culture, region, nation, population, and ethnicity that has learned about it has adopted from Europe. The art, music, literature, architecture, economic order, legal codes, ethical and political value systems of the West have not secured universal acceptance. Indeed, once decolonialization set in, these "blessings" of European culture have more often than not been repudiated by non-Europeans. But not science. And we need not say "Western" science. For there is no other kind, nor did science really spring up independently elsewhere before, simultaneously with, or after its emergence among the Greeks 2,500 years ago. It is true that some technologies that facilitated Western political, military, and economic dominance over much of the rest of the world, like gunpowder, moveable type, and pasta, originated elsewhere, principally in China. And several non-Western civilizations kept substantial and detailed records of celestial phenomena. But isolated technological achievement and astronomical almanacs are not science; the predictive powers that accompanied these achievements were not harnessed to an institutional drive to improve rational understanding that is characteristic of Western science from the ancient Greeks through medieval Islam and Renaissance Italy to the Protestant Reformation and twentieth-century secularism.

The emergence of science solely in the West, and the universality of its embrace across all the non-Western civilizations, raises two distinct questions. First, why only at first in the West? Second, what is it about science that led to its adoption by cultures not otherwise interested in distinctively Western ideas, values, or institutions?

To the first question some answers can be immediately ruled out. Neither the ancient Greeks among whom theoretical science emerged, nor the Muslim cultures by whom it was preserved, nor for that matter the Renaissance

Europeans who so rapidly accelerated its development are, as peoples, intellec-
tually more able or naturally more curious than any other peoples around the
world. Science is not "in their genes, in their DNA." Nor is it reasonable to credit
the emergence of science, its preservation or its flourishing to any one or a small
number of individuals, say Euclid, Archimedes, Avicenna, Galileo, or Newton.
The achievements of one or a small number of individuals are all too likely to be
swamped by the indifference of the many. Besides, it is more than likely that soci-
eties from pre-Christian Mesoamerica to latter-day New Guinea have produced
individuals equal in their special gifts to these path-breaking scientists.

One attractive explanation for the origination of science in the West owes
a good deal to a book by Jared Diamond, *Guns, Germs, and Steel*. Diamond
did not set out to explain why science emerged in the West and why it came
eventually to be adopted everywhere, even in cultures inimical to other Western
institutions. Rather Diamond wanted to explain European political and mili-
tary domination of so much of the rest of the world up to the twentieth cen-
tury. And his starting point was the relative equality in intelligence and cultural
attainments of all *Homo sapiens* everywhere when the hunter-gatherer mode
of existence ceased to be an adaptive response to the local environment almost
everywhere 10,000 years ago.

Diamond marshals a great deal of evidence to show that Western Europe
did not become the dominant force in the world owing to the superiority of its
institutions, culture, or civilization, still less some individual difference between
Western and non-Western peoples. Instead the West's success in colonizing, sub-
jugating, and exploiting the rest of the world was the result of a small number of
very "natural" geographic and environmental factors. First, of the dozen or so
species of easily and profitably domesticable plants, half grow in one region: the
Near East. Accordingly, agriculture could be expected to begin there. With agri-
culture comes storable goods and the need for record keeping, so writing began
earliest there as well (and began independently in Mesoamerica approximately
1,000 years later from the same cause, the domestication of storable corn and the
consequent need for record keeping). Agricultural productivity is enhanced by
the domestication of traction (pulling) animals. However, of the 18 or so poten-
tially domesticable traction animals, again the plurality are to be found in the
Near East. In some regions where domesticable plants occur (e.g. Mesoamerica),
there are no indigenous animals to domesticate for traction (pulling plows, sleds,
wagons). Higher agricultural production increases population, and in dense
populations domesticated animals communicate epidemic diseases to people.
This causes near-term population collapse but long-term disease resistance.
So after many generations, almost the entire remaining population is immune
to these originally animal-borne diseases. Thus, Near Eastern populations
equipped with storable foodstuffs, and effective (traction) transportation, were
able to respond to population pressures by expansion into occupied and unoccu-
pied territories (initially of Europe) far from their places of origin.

Diamond makes another crucial observation: there are no geographic or
climatic barriers to the lines of communication along which technological

innovations can move, all the way from the Atlantic coast of Europe to the Far East Pacific along the band between 30 and 45 degrees north latitude. By contrast, the lines of communication between any two points in North and South America must find a way across the very narrow, very mountainous, and mosquito-infested Isthmus of Panama. Similarly the path of transmission of technological innovation in Africa is broken by the Sahara and the insect-borne disease regions immediately south of it. Accordingly, the access of peoples anywhere along the Eurasian axis to novel technologies is far greater than those of the western hemisphere, Oceania, or Africa. Finally, the European continent is itself characterized by a large number of interior mountain barriers and a coastline crenulated by potential harbors, with rich fisheries just beyond the sight of land. These environmental factors selected for relatively early expertise in beyond-sight-of-land sailing.

Altogether, the natural agricultural and traction-animal advantages of Near Eastern and European peoples, their early acquisition of immunities to animal-borne diseases, together with long-term access to technological innovation from as far away as China and Japan, and the relatively greater environmental incentives to ocean navigation, make it pretty much inevitable that Western European populations would arrive on distant shores. There the diseases they or their animals carried would be likely to kill substantial proportions of local inhabitants before the local population could return the favor. After that it was weapons and transport that enabled Europeans to dominate the survivors. This outcome is, from the perspective of the twenty-first century, by no means a good thing. Indeed, it was a very bad thing in terms of the human and cultural loss to their victims and the moral harm that European occupiers brought upon themselves. Here Diamond's theory stops. But it is relatively easy to extend it to explain the emergence of science first in the West.

That pure science should have been expected to emerge earliest among the more technologically sophisticated societies is not a claim Diamond makes, but it is a fairly obvious inference to draw from his analysis. After all, the difference between enquiries in engineering and in pure science is clearly a matter of degree, and the serendipitous course of inquiry is bound to lead from technology to pure science if given enough time. It is inevitable that the search for practical improvements in technology should at least sometimes lead to explorations in pure as opposed to applied science. Thus, the earlier the onslaught in a society of "guns and steel," if not germs, the earlier that what we recognize as science comes to flourish in that society. That is why science emerged earliest in the West.

Why Is Science the Only Feature of Western Culture Universally Adopted?

Let's turn to the second of our two questions: Why is science the sole distinctively Western achievement to have been adopted by every other culture on the planet capable of doing so? It would at first blush appear that the explanation sketched above for why science emerged initially in the West would also provide

an answer to our second question: Once science is available, individuals and societies everywhere will seek the sort of objective knowledge of reality that pure science has provided in the West. So, individuals and groups everywhere will adopt the methods of science. This simple extension of our explanation makes several mistakes, some of them subtle. First, the explanation for why science should emerge initially in the West identifies **necessary conditions** for its emergence that obtained earliest in the West, not the **sufficient conditions** that obtain and would explain its adoption everywhere else. Second, for all we know, besides the sufficient conditions that obtained first in the West, there may be other conditions, cultural values, social practices, political institutions, economic conditions that obstruct the discovery and adoption of scientific methods present in non-Western cultures. If there are such further conditions, then science has established itself in these non-Western societies by overcoming, changing, or otherwise trumping the indigenous values, practices, institutions, and conditions of these peoples. Third, the explanation presumes that other cultures share the West's interests in technological improvement. Some of them may not do so. Fourth, and perhaps most surprising to those unacquainted with controversies surrounding science, the assumption that Western science has been characterized by persistent improvements in objective knowledge of the world has been widely challenged by historians and sociologists of science and other postmodern thinkers (see Chapter 14). The question of why science spread so quickly and uniformly still needs an answer. The first steps toward one require an understanding of science itself.

Our second question—about why science is universally adopted—thus remains open. It will be especially acute if we identify standards of objective knowledge associated with science not shared by or even rejected by other cultures. The practice of scientific inquiry is widely supposed to require disinterestedness and the rejection of authority, and to institutionalize skepticism and prohibit the ownership of ideas, requiring that data and methods be public and equally shared. These requirements are at variance with the mores of many non-Western cultures (and more than a few Western governments in the last century). If science embodies such standards, values, methods, and practices, whether the mores of non-Western societies would impede its universal adoption turns out to be an important matter. And if they clash with the values of non-Western cultures, then explaining how and why they have won out in competition with them will require further inquiry. Finally, if the methods of science were not originally adopted in the West owing to the objective knowledge of nature they now provide, as not a few influential scholars have sought to show, then not only will our second question remain open, but the answer to our first one, why science emerged first in the West, may have to be rejected.

Quite independent of their intrinsic interest, these issues make understanding what science is, how it works, what its methods, foundations, values and presuppositions are a pressing matter. These are tasks that the philosophy of science long ago set itself.

In the last 50 years or so, philosophy has been joined in its absorption in these issues by other disciplines such as the sociology, psychology, and economics of science and other social and behavioral studies of science. These disciplines have burgeoned in the last three decades, and there are now large numbers of psychologists, sociologists, and other students of science eager to enhance our understanding of science. How do the interests of the philosophy of science differ from the agenda of these late twentieth-century disciplines? Can it claim to have some priority over these disciplines in the quest for an understanding of science? Here are some tentative answers to these questions.

To begin with, the other enterprises—the sociology, psychology, economics and politics of science—are themselves presumably scientific ones: to the extent possible, they hope to share the methods of science in their own inquiries into the social, psychological, economic, political characteristics of science. But until we are clear about what the methods of science are, these enterprises are at risk of frustration and failure in attempting to attain their scientific objectives, for they will be unclear about the means to reach their scientific goals. This does not mean that we cannot do science of any kind until we have established what exactly the methods of science are, and ascertained their justification. But it does mean that we should scrutinize those sciences already widely recognized as successful in the pursuit of their objectives, in order to identify the methods likely to succeed in less well-developed sciences, such as the sociology or psychology of science.

But this scrutiny cannot be sociological, psychological, economic, or political, at least not at the outset. For science as a product or result—the concepts, laws, theories, methods of experiment, and observation—is not supposed to reflect or even allow for the operation of factors studied in these disciplines like sociology or psychology, economics, politics or history, or of social status, personality types, obvious financial incentives, political power or cognizance of historical precedent. The considerations that appear to drive scientists' discussions and debates, and their acceptance and rejection of findings and theories, call upon notions of logical reasoning, evidence, testing, justification, explanation, with which philosophy has grappled since Plato. It is possible that, in the end, analysis of and reflection on these notions and how they operate in science cannot answer our questions about its character, nor sanction its claims to provide objective knowledge that other enterprises seek to secure. If this turns out to be the case, then we may usefully turn to the social and behavioral studies of the nature of science for real elucidation of the value of this distinctive contribution of the West to world civilization. Yet first we have to wrestle with the philosophy of science.

Summary

The special place of science as a source of objective knowledge raises questions about how it secures such knowledge and whether there are alternative sources

or means of securing it. Because it has always provided an influential description of reality, science has historically been the strongest force in shaping the pressing philosophical problems. Indeed, some philosophical problems track changes in natural science. How philosophers think about the mind and its place in nature, free will versus determinism, the meaning of life, all are deeply affected by scientific developments. As science's descriptions of reality have changed over the centuries, the philosophical problems have changed as well.

Since science is arguably the only distinctive feature of Western civilization that all the rest of the world has taken up, understanding science is an important part of coming to grips with the influence—whether good or bad—that the West has had on other cultures. Answering this question requires that we understand what science is. Philosophy has a better claim than other disciplines to be allowed to give an initial answer to the question of what science consists in.

Study Questions

1. Given the amount of change in the scientific conception of the world over the centuries, does philosophy pay too much attention to its findings and theories in dealing with philosophical problems?
2. Defend or criticize: "Philosophy is much more difficult than science, even though there is no mathematics and no labs."
3. Defend or criticize: "If there is no litmus test that will tell science from pseudoscience, there is no difference between them at all."
4. "As an open-minded and objective inquiry into the nature of the world, science should welcome the sort of unorthodox research which an agency like the Office of Alternative Medicine is designed to encourage." Are there good grounds for this claim?
5. Defend or criticize: "The claim that science is a uniquely Western contribution to the world is ethnocentric, uninformed, and irrelevant to understanding science's character."
6. Does the philosophy of science's conception of the nature of science compete with the sociology of science's conception of its nature?
7. Minorities and women were long excluded from science. They also played a limited role in the philosophy of science. Did their exclusion undermine the reliability of either one of these distinct undertakings?

Suggested Readings

Important natural scientists have always extrapolated from their own scientific achievements to philosophical conclusions, i.e. answers to questions science cannot yet (or perhaps can never) answer. Among these perhaps the most important was Albert Einstein, much of whose reflection on philosophy of

science (and other compartments of philosophy) was subject to philosophers' scrutiny, and Einstein's own reflection on the philosophers' scrutiny, in P. A. Schilpp, *Albert Einstein: Philosopher-Scientist*. More recent (mostly critical) commentary on philosophical works by physicists include Richard Feynman, *The Character of Physical Law*, and Steven Weinberg, *Dreams of a Final Theory*. Among biologists, the same temptation has produced E. O. Wilson, *Consilience*, a sustained argument for the thesis that natural science can answer all but the pseudo-questions, and R. Levins and R. Lewontin, *The Dialectical Biologist*, which adopts a view quite the contrary to Wilson's.

Richard Dawkins, *The Blind Watchmaker*, is an excellent introduction to Darwinism and the theory of natural selection, but it is no substitute for reading Charles Darwin, *On the Origin of Species*. The best introduction to the mysteries of quantum theory for the non-specialist is Bryan Greene, *The Elegant Universe*. Adam Becker, *What is Real? The Unfinished Quest for the Meaning of Quantum Physics*, sketches the twentieth-century history of this subject and its vexed relationship to the philosophy of science.

Tim Maudlin, *The Metaphysics within Physics*, is an exploration of philosophical problems that physics cannot avoid.

E. Nagel and J. R. Newman, *Gödel's Proof*, provides an accessible account of this central mathematical result.

Paul Thagard, "Why Astrology Is a Pseudoscience" and Michael Ruse, "Creation Science Is Not Science" are reprinted in Martin Curd and J. A. Cover's *Philosophy of Science: The Central Issues*, as is a critical discussion of uncritical demarcation tests by Larry Laudan, "Commentary: Science at the Bar—Causes for Concern," with a response by Ruse.

Lee McIntyre, *The Scientific Attitude: Defending Science from Denial, Fraud, and Pseudoscience*, is a contemporary exploration of how scientific values can be used to push back against science denial and pseudoscience.

Important works in sociology of science start with R. Merton, *The Sociology of Science*. A view about the relation of sociology and philosophy of science quite different from those advanced here can be found in D. Bloor, *Knowledge and Social Imagery*. B. Barnes, D. Bloor, and J. Henry, *Scientific Knowledge: A Sociological Analysis*, offers a revision of his earlier strong opposition. A. Pickering, *Constructing Quarks*, applies a sociological analysis to account for scientific discovery. See also the Suggested Readings in Chapter 14.

3   Scientific Explanation

Overview

Science, like other human activities, is one response to our need to understand the world. The way it does so differs, though, from possibly competing activities like religion, mythology, or for that matter common sense. Science claims to provide objective explanations that are superior to other approaches when dealing with empirical phenomena.

Different accounts of what it means for science to "explain" reflect fundamental philosophical differences that go back to Plato—between those who view scientific explanation as something we discover versus those who treat it as something we create. The logical positivists aimed to formulate an ideal standard of explanation for scientists to aspire to. Other philosophers have sought to understand scientific reasoning through the explanations that scientists actually give.

One starting point for understanding scientific explanation focuses on the role of laws of nature. **Scientific laws** have explanatory force presumably because they describe the way that things have to be. But the way things have to be—the necessity of laws of nature—is very difficult to understand from the scientific point of view, for scientific observation and experiment never show how things have to be, only how they are.

Dissatisfaction with logical positivists' answers to this question shifted the focus of some philosophers of science away from laws as explanatory. This approach leads to a theory of explanation that focuses on how explanations answer people's questions, instead of what ingredients they must have to be scientific.

Defining Scientific Explanation

Philosophy, the logical positivists recognized, cannot provide explanations that compete with those of science. What it could provide was an "explicit definition" or a "**rational reconstruction**" or what would now be called a "conceptual analysis" of scientific explanation. Though such an analysis would give the meaning of the concept of explanation, it would be more than a dictionary definition. A dictionary definition merely reports how scientists and others actually use the words "scientific explanation." Positivists, and the philosophers of science who continue their tradition of philosophical analysis, seek a check-list of conditions that any scientific explanation should satisfy. When all are satisfied, the checklist guarantees the scientific adequacy of an explanation. In other words, the traditional approach seeks a set of individually necessary and jointly sufficient conditions for something to be a scientific explanation. This "rational reconstruction" of the dictionary definition would render the concept of scientific explanation precise and logically clear.

An explicit definition gives the necessary and sufficient conditions for a thing, event, state, process or property to be an instance of the term defined. For example: "triangle" is explicitly defined as "plane figure having three sides." Since the conditions are together sufficient we know that everything that satisfies them all is a Euclidean triangle and since the conditions are individually necessary, we know if just one is not satisfied by an item, it is not a Euclidean triangle. The beauty of such definitions is that they remove any vagueness, and provide maximally precise definitions.

An explicit definition or "**explication**" of the notion of a scientific explanation could serve the prescriptive task of a litmus test or yard-stick for grading and improving explanations in the direction of increasing scientific adequacy. The demand that philosophical analysis result in such a precise and complete definition is in part a reflection of the influence of mathematical logic on the logical positivists and their immediate successors in the philosophy of science. For in mathematics concepts are introduced in just this way—by providing explicit definitions in terms of already understood, previously introduced concepts. The advantage of such definitions is clarity: there will be no borderline cases and no irresolvable arguments about whether some proposed explanation is "scientific" or not. The disadvantage is that it is often impossible to give such a complete definition or "explication" for concepts of interest.

Let's call the sentences in an explanation that do the explaining the "**explanans**" and those that report the event to be explained the "**explanandum.**" There are no

convenient single English words for these terms, so they have become common-place in philosophy. An examination of the kinds of explanations that almost all scientists find acceptable makes it pretty obvious that scientific explanans usually contain laws: when the explanandum is a particular event, like the Chernobyl reactor accident or the appearance of Halley's comet in the night sky over western Europe in the fall of 1986, the explanans will also require some "**initial**" or "**boundary conditions**." These will be a description of the relevant factors—say the position and momentum of Halley's comet the last time it was sighted, or the position of the control rods of the reactor just before it overheated—which together with the law result in the explanandum-event. In the case of the explanation of a general law, like the ideal gas law

$$PV = nRT,$$

the explanans will not contain boundary or initial conditions. Rather it will contain other laws, which work together to explain why this one obtains. In the case of the ideal gas law the other laws that work together to explain it are the parts of the kinetic theory of gases.

Suppose we want to know why the sky is blue, a question people have asked probably as far back as human origins. Now this is a particular state of affairs at a particular place, the Earth. The Martian sky presumably is reddish in hue. So, to explain why the sky on Earth is blue we require some information about "boundary conditions" and one or more laws. The relevant boundary conditions include the fact that the Earth's atmosphere is composed of molecules mainly of nitrogen and oxygen. It's a scientific law that gas molecules scatter the light that strikes them in accordance with a mathematical equation first formulated by the British physicist Rayleigh. The amount of light of any wavelength scattered by a gas molecule depends on its "scattering coefficient"—$1/\lambda^4$—one over its wavelength to the fourth power. Since the wavelength of blue light is 400 nanometers (another law), and the wavelength of other light is greater (for example, red light has a wavelength of 640 nanometers), the scattering coefficient of nitrogen and oxygen for blue light is greater than for other light. Therefore, the molecules in the Earth's atmosphere will scatter more blue light than other colors, and the atmosphere will look blue. In a physics text this explanation is laid out in more detail, with the relevant equations derived and the amounts of scatter calculated.

Examples from the social and behavioral sciences are easier to understand because they are less quantitative. But explanations in social science that everyone accepts are harder to come by in these disciplines because we have discovered few if any laws in them. Thus, some economists will explain why the rate of interest is always positive (a general "law") by deriving it from other general laws, such as the "law" that, other things being equal, people prefer immediate and certain consumption to future and uncertain consumption. From this law it follows that to get people to defer consumption to the future, you have to pay them by promising that they will have more to consume later if they postpone

consumption, and instead invest what they would have consumed to produce more. The payment for postponed consumption is measured as the interest rate. As in physics, the explanation here proceeds by derivation, this time of a law (instead of a particular fact, from other laws. Here we don't need boundary conditions because we are not explaining a particular fact). But the explanation still employs laws, if, that is, these generalizations about people are indeed laws. (Some economists reject this explanation for why interest rates are always positive. They hold that other factors besides preference for immediate consumption explain this generalization. In effect, they reject the general claim that the interest rate reflects the price of people's willingness to postpone consumption is a law, or even a true statement.)

The Role of Laws in Scientific Explanation

It's obvious that scientific explanations contain laws or hypotheses or generalizations of some sort on which they at least implicitly rely. But must they do so? And if so, why? Positivists and later philosophers of science could not be satisfied merely reporting this fact about the practice of science. They needed to account for why laws are present and why according to their rational reconstruction of scientific explanation, laws seemed indispensable.

Why must a scientific explanation contain one or more laws? What is it about laws that is explanatory? This question is the main topic of the next chapter. But we can immediately sketch one reason that is widely given. This answer to the question begins with the claim that scientific explanation is just causal explanation. Scientists search for causes. They do so because science seeks explanations that also enable us to control and predict phenomena, and this is something that only knowledge of causes can provide. If scientific explanation is causal explanation, then by a well-known philosophical theory of **causation**, it must explicitly contain or implicitly assume laws. The traditional empiricist account of causation that goes back to David Hume in the eighteenth century holds that the relation of cause and effect obtains only when one or more laws *subsume* the events so related—that is, cover them as cases or instances of the operation of the law. Thus, the initial or boundary conditions of the explanans cite the cause of the explanandum phenomenon, which are the effects of the boundary conditions according to the law mentioned in the explanans.

On the empiricist view, causation consists in law-governed sequence because there is no other observationally detectable property common to and distinctive of all causal sequences besides exemplifying general laws. When we examine a single causal sequence—say one billiard ball hitting another, and the subsequent motion of the second ball—there is nothing to be seen that is not also present in a purely coincidental sequence, like a soccer goalkeeper's wearing green gloves and her successfully blocking a shot. The difference between the billiard-ball sequence and the green goalie-glove sequence is that the former is an instance of an oft-repeated sequence, and the latter is not. The last time the goalie wore the green gloves she failed to stop the shot.

All causal sequences share one thing in common that is missing in all merely coincidental sequences: they are instances of—they instantiate—general laws. This philosophical theory does not require that for every causal claim we make we already know the law or laws that connect the cause and effect. Children will explain, correctly we suppose, why the vase broke, by admitting that it was dropped (passive voice, silence on who dropped it), on a marble floor. We accept the statement as identifying the cause, even though neither the children nor we know the relevant laws. The theory that causal sequences are instances of laws of nature doesn't require that we accept the statement. It only requires that there is a law or laws, already known or not yet discovered, which do so. The task of science is to uncover these laws, and to employ them in the explanations of effects. If scientific explanation is causal explanation, and causation is law-governed sequence, then it follows fairly directly that scientific explanations require laws.

But there are problems with this argument for why scientific explanations must cite laws. First of all, a few important examples of scientific explanations don't cite causes, or don't do so in any obvious way. Consider the ideal gas law, for example:

$$PV = nRT.$$

This law explains the temperature of a gas at equilibrium by appeal to its pressure and the volume it takes up. But volume and pressure can't be causes of temperature since all three—the temperature, the volume, and the pressure—vary, in the way the law describes, instantaneously. It's simply not the case that changes in volume at one time cause changes in temperature at a later time; the change in temperature occurs during exactly the same interval that pressure is changing. But as both a good deal of philosophy and contemporary physics require, causes have to precede their effects.

There is a second and perhaps more formidable objection to the claim that scientific explanations require causal laws, which is that instead of making things clearer, it threatens to involve the analysis of scientific explanation in a thicket of "metaphysical" issues that the positivists and other philosophers sought to avoid. These are issues about the nature of causation that go back in philosophy all the way through the rationalists and the empiricists to Plato and Aristotle!

The nature of causation has been controversial for millennia. There is certainly no consensus on Hume's claim that every causal sequence is causal just because it is law-governed. Many philosophers have held that causation is a much stronger relation between events than mere regular succession. Thus, the sound of thunder regularly succeeds the flash of lightning, but the latter is not its cause. Rather they are joint effects of a common cause, the electrical discharge from the cloud to the earth. Most philosophers have agreed that causes somehow *necessitate* their effects and that mere regularity cannot express this necessity. The logical positivists who first advanced an explicit account of scientific explanation wished strongly to avoid traditional controversies about the existence and nature of causal necessity. Ever since Aristotle, philosophers and scientists had engaged in fruitless discussion and debate about the kinds of causes there are,

the nature of causation, the kind of necessity it embodied, how widespread was its domain—did it include biology, human life, thought, and action—and so on. Such questions were deemed "metaphysical" by the positivists because no scientific experiment could answer them. But, as we shall see in the next chapter and elsewhere in this book, once philosophy gave up positivism, it and science could not avoid these and other metaphysical problems.

In addition to the interest in avoiding metaphysical issues, some empiricist philosophers held that the notion of causation was an obsolete, anthropomorphic one, with misleading overtones of human agency, manipulation, or power over things. Accordingly, these philosophers needed a different argument for the requirement that scientific explanations must contain laws in their explanans.

The argument that logical positivists advanced for the role of laws in explanations illuminates several aspects of their philosophy of science. To begin with, these philosophers sought a notion of scientific explanation that would constitute an objective relationship between explanandum and explanans, a relationship like the relation of mathematical proof, which obtains regardless of whether anyone recognizes it or not; a relationship that is sufficiently precise that we can determine whether it obtains without any doubt or borderline cases. Thus, the logical positivists rejected the notion of scientific explanation as an attempt to allay curiosity or answer a question that might be put by an inquirer, which seemed overly contingent. It is relatively easy to "explain" complex physical processes to children by telling them stories that allay their curiosity. The subjective psychological relevance of the explanans to the explanandum in such cases may be great, but they do not constitute scientific explanations. The logical positivists were not interested in examining how a scientific explanation might be better or worse, appropriate or inappropriate, given the beliefs and interests of someone who might ask for the explanation. The conception of explanation as an answer to someone's question is not one these philosophers sought to explicate. Rather they sought an explication or "rational reconstruction" of the concept of explanation that would provide the sort of role in science that the notion of "proof" plays in mathematics. The problem of explanation for logical positivists was to find some conditions on explanation that ensured the objective relevance of the explanans to the explanandum. They needed a relationship that made explanatory relevance a matter of objective relations between statements and not the subjective beliefs about relevance of less-than-omniscient cognitive agents. Only by invoking laws could explanations identify the objective relationship between explanans and explanandum. Let us now see exactly why this is so.

The Covering Law Model

The logical positivists' model of scientific explanation is based on the idea that the explanans is relevant to the explanandum in that the explanans gives good grounds to have expected that the explanandum-event would have happened. You may be surprised by this requirement. After all, when we ask for the explanation of an event, we already know that it has happened. But satisfying this

requirement involves producing further information which, had we been in possession of it before the explanandum-event occurred, we would have been able to predict it. Now, what kind of information would allow us to satisfy this requirement? A law and a statement of boundary or initial conditions will enable us to fulfill this requirement if the law and the boundary conditions together logically imply the explanandum. The relation of logical implication has two important features. First it is truth-preserving: if the premises of a **deductively valid argument** are true, then the conclusion must also be true; second, whether the premises of an argument logically imply the conclusion is an objective matter of fact that can in principle be decided mechanically (for example, by a computer). These features fulfill the demand that the logical positivist made of an explication of the concept of scientific explanation.

This analysis of scientific explanation, associated most closely with Carl G. Hempel, the philosopher who did the most to expound and defend it, came to be called the "**deductive-nomological (D-N) model**" ("nomological" from the Greek *nomos* meaning lawful). Critics of this D-N account of explanation labeled it (and its statistical extensions) the "**covering law model**" and this name came to be adopted by its defenders as well. Hempel's fundamental idea was the requirement mentioned above—that the explanans give good grounds to suppose that the explanandum phenomenon actually occurred. It stands as his "general adequacy criterion" on scientific explanations.

In Hempel's original version the requirements on deductive nomological explanation were as follows:

1. The explanation must be a valid deductive argument.
2. The explanans must contain at least one general law actually needed in the deduction.
3. The explanans must be empirically testable.
4. The sentences in the explanans must be true.

Between them, these four conditions are supposed to be individually necessary and jointly sufficient for any set of statements to constitute a scientific explanation of a particular fact. Notice that an explanation that satisfies these conditions provides enough information so that one could have predicted the occurrence of the explanandum-event, or similar events, given knowledge that the initial or boundary conditions obtain. Thus, the D-N model is committed to in principle symmetry between explanation and prediction. In fact, this commitment already follows from the objective relevance requirement stated above.

The first condition guarantees the relevance of the explanans to the explanandum. The second is needed so as to exclude patently non-explanatory argument like:

1. All free-falling bodies have constant acceleration.
2. It rained on Monday.
   Therefore,
3. It rained on Monday.

Notice that this argument satisfies all of the other conditions on explanation. In particular, it is a deductively valid argument because every proposition deductively implies itself, so premise 2 implies premise 3 But it is no explanation, if only because nothing can explain itself! And of course it's not a D-N explanation for another reason: the law it includes is not needed to make the deduction valid. Consider another example.

1. All puppies born in this litter have a brown spot on their foreheads.
2. Fido is a puppy born in this litter.
     Therefore,
3. Fido has a brown spot on his forehead.

This argument is no explanation of its conclusion owing to the fact that premise 1 is no law of nature. It's an accident of genetic recombination at best.

The third condition, **testability**, is supposed to exclude non-scientific explanations that make reference to explanatory factors that cannot be subject to confirmation or disconfirmation by observation, experiment, or other empirical data. It reflects the epistemological commitment of empiricism about scientific knowledge: The requirement that the explanans be testable is meant to exclude non-scientific and pseudo-scientific explanations, such as those offered by astrologers, for example. How testability is assured is a subject to which we turn in Chapter 10.

The fourth condition, that the explanans must be true, is problematical and introduces some fundamental philosophical problems, indeed the very ones the logical positivists hoped to escape by silence about causation. Every scientific explanation must include a law. But laws are by definition true everywhere and always—in the past, in the present, in the future, here and everywhere else in the universe. As such, they make claims that cannot be established conclusively. After all, right now we have no access to the distant past or even the nearest future, let alone all places and times where events happen that make laws true. That means that the statements we believe to be laws are at best hypotheses that we cannot know for sure. For convenience let's distinguish between "**natural laws**," true everywhere and always whether we have uncovered them or not, and "**scientific laws**," which is what we will call those hypotheses well established in science as our best current estimates of what the natural laws are.

Since we cannot know whether our scientific laws are natural laws, that is, whether they are true, we cannot ever know for sure that any explanation satisfies condition 4 above: that the explanans be true. Indeed, the situation is worse: since every previous hypothesis we have advanced about the natural laws has proved to be wrong, and been replaced by a more accurate scientific law, we have excellent reason to suppose that our current scientific laws (our current best guesses about what the natural laws are) are wrong too. In that case, we have equally good reason to think that none of our current scientific explanations really satisfy the deductive nomological model.

But what's the use of an analysis of explanation according to which we probably have never uncovered any scientific explanations, only at most approximations to them, whose degree of approximation we can never measure?

We might try to avoid this problem by weakening requirement 4. Instead of requiring that the explanans be true, we might require that the explanans be our best current guesses about the natural laws. The trouble with this weakened requirement is twofold. It is by no means clear and precise which are our best guesses about natural laws. Physicists disagree just as social scientists do about which guess is the best one, and philosophers of science have by no means solved the problem of how to choose among competing hypotheses. In fact, the more one considers this question the more problematical becomes the nature of science, as we shall see in Chapters 11 and 12. Weakening the requirement of truth into the requirement that the explanans include the most well-established currently known scientific law (i.e. our best guess hypothesis) thus undermines the D-N model's claims to precision in explication.

The second problem we face is the nature of scientific laws and natural ones. Two of our four conditions on a scientific explanation invoke the notion of a law. And it is pretty clear that the explanatory power of an explanation in the natural sciences is in fact borne by the law. This is something almost universally accepted by even those who reject the covering law model of explanation. The scientific law is what makes explanatory the connection between the particular facts mentioned in the initial conditions of the explanans, and the particular facts mentioned in the explanandum. We must therefore face the question: What exactly is a natural law and what is it about such laws that makes them explanatory?

Problems for the Covering Law Model

Progress in the philosophy of science has often consisted in the construction of counterexamples to analyses, definitions, or explications, and then revisions of the definition to accommodate the counterexamples. Since the sort of analysis traditionally preferred by logical positivists provides a definition in terms of individually necessary and jointly sufficient conditions for the concept being explicated, counterexamples can come in two different forms: first, examples that most informed persons will concede to be explanations but which fail to satisfy one or more of the conditions laid down; second, an example that no one takes to be an acceptable scientific explanation, but which by hook or by crook somehow satisfies all conditions for being one.

Counterexamples to the D-N model of the first sort have often been found in history and the social sciences, where the most well-accepted explanations often fail to satisfy more than one of the D-N model's conditions, especially the requirement that laws be cited. For example, the explanation of why Britain entered World War I against Germany does not seem to involve any laws. Imagine someone framing a law such as "Whenever Belgian neutrality is protected by

treaty and is violated, then the treaty signatories declare war on the violator." Even if the proposition were true, it's no law, not least because it names a specific place in the universe. If we substitute for "Belgian" something more general, such as "any nation's," the result is more general, but plainly false. One response to the fact that many explanations don't cite laws that is often made in defense of D-N explanation is to argue that such explanations are "explanation-sketches" that could eventually be filled out to satisfy D-N strictures, especially once we have uncovered all of the boundary conditions and relevant laws of human action. Counterexamples of this sort in the natural sciences are more difficult to find, and defenders of the D-N model are confident they can deal with such cases by arguing that the alleged counterexample does satisfy all conditions. Thus, consider the explanation of the *Titanic*'s sinking. Her sinking was caused by collision with an iceberg. Surely this explanation will be accepted even though there is no law about the *Titanic*, nor even one about ships that strike icebergs sinking. The explanation is an acceptable one even when we note that it is often offered and accepted by persons who know almost nothing about the tensile strength of iron, the coefficient of elasticity of ice, or the boundary conditions which obtained on the night of April 14, 1912 in the North Atlantic. Presumably, a naval engineer could cite the relevant laws along with the boundary conditions—size of the iceberg, speed of the *Titanic*, composition of its hull, placement of its water-tight doors, etc., which underlie the explanation-sketch and enable us to turn it into a D-N explanation.

   Counterexamples of the second sort, which challenge the sufficiency of the D-N conditions as a guarantee of explanatory adequacy, are more serious. Among the most well-known is the "flag-pole's shadow" counterexample due originally to Sylvain Bromberger. Consider the following "explanation" for the fact that at 3:00 p.m. on July 4, 2000, the flag-pole at City Hall in Missoula, Montana is 50 feet high:

1. Light travels in straight lines.
2. At 3:00 p.m. on July 4, 2000 the sun is emitting light at a 45-degree angle to the ground where the flag-pole is located, perpendicular to the ground. (boundary condition)
3. The shadow cast by the flag-pole is 50 feet long. (boundary condition)
4. A triangle with two equal angles is isosceles. (mathematical truth)
    Therefore,
5. The flag-pole is 50 feet high.

The "explanation" is designed to satisfy all four conditions given for D-N explanations above, without being a satisfactory explanation of the height of the flag-pole. Why? The deductive argument fails to be an explanation presumably because it cites an effect of the flag-pole's height—the shadow it casts—not its cause, which might be the desires of the Missoula city mothers to have a flag-pole one foot taller than the 49-foot flag-pole at Helena, Montana.

One conclusion sometimes drawn from this counterexample is simply to reject the whole enterprise of seeking an objective explanatory relation between statements about facts in the world independent of the human contexts in which explanations are requested and provided. To see why such a move might be attractive, consider whether we could construct a context in which the deduction above is in fact an acceptable explanation for the height of the flag-pole. For example, suppose that the city mothers had wished to build the flag-pole to commemorate the American commitment to equality and union by casting a shadow exactly equal in length to the pole and exactly as many feet as there are states in the union at the moment annually chosen for patriotic exercises on American Independence Day. In that case, Bas van Fraassen has argued, for someone well informed about the wishes of the city mothers, it would be a correct answer to the question "Why is the flag-pole 50 feet high?" to reply in the terms mentioned in the deductive argument above.

This argument is supposed to show that explanation is not merely a matter of logic and meaning (syntax and semantics), but is as much a matter of "**pragmatics**" (that dimension of language which reflects the practical circumstances in which we put it to use). We may contrast three different aspects of a language: its syntax, which includes the rules of logic as well as grammar; its semantics, the meanings of its words; and its pragmatics, which includes the conditions that make some statements appropriate or meaningful. For example, it's a matter of the pragmatics of language that "Answer yes or no, have you stopped beating your dog?" is a question we can only ask dog-beaters. A man who does not own a dog or one not given to dog-beating cannot answer this question with a yes or a no. Similarly, if explanation has a pragmatic element we can't tell when something successfully explains unless we understand the human context in which the explanation was offered.

The pragmatics of language is presumably something we can ignore in mathematical proof, but not, it is argued, in scientific explanation. Whether an analysis of scientific explanation must include this pragmatic dimension is a topic for the next section. But one point that can be made now is that even if explanation is unavoidably pragmatic, it may still turn out that the D-N model provides important necessary conditions for scientific explanation—to which some pragmatic conditions need be added. Indeed, it may be that the D-N model provides the distinctive features of scientific explanation, while the pragmatic element provides the features common to scientific and non-scientific explanations.

Another implication sometimes drawn from the flag-pole counterexample is that the D-N model is inadequate in not restricting scientific explanations to causal ones, or at least in not excluding from the explanans those factors later in time than the explanandum. Notice that the casting of a shadow 50 feet long at 3:00 p.m. on July 4 is something that happens well after the flag-pole was first fabricated at 50 feet in height or mounted vertically. But what is the reason for this restriction? Evidently it is our belief that causation works forward in time, or at least not backwards, and that somehow the direction of explanation must follow the direction of causation. So, we might add to the D-N model the

additional condition that the boundary conditions be the prior causes of the explanandum. The trouble with this addition to our requirements on explanation is that there appear to be scientific explanations that do not invoke temporally prior causes. Recall, for example, how we explain the temperature of a gas at equilibrium in terms of the ideal gas law, $PV = nRT$ and the boundary condition of its simultaneous pressure and the volume of the vessel in which it is contained. Simultaneous levels of pressure, volume, and temperature do not cause one another. So this D-N explanation is not obviously causal.

Worse still, this addition invokes causation to preserve the D-N model, and causation is something about which the proponents of D-N explanation wanted to remain silent. Although the logical positivists tried, philosophers of science were eventually unable to continue to maintain a dignified silence about the embarrassingly metaphysical problems of causation owing to another obligation they bore: that of providing an account of how statistical explanation works.

Both the social and biological sciences have long been limited in providing such explanations because they have not uncovered universal non-statistical laws. And the indeterminacy of sub-atomic physics makes such explanations arguably unavoidable, no matter how much we learn about nature. It may seem a straightforward matter to extend the D-N model to statistical explanations. But it turns out that this straightforward extension is another reason to take the pragmatics of explanation seriously, or at least to treat explanation as a relation between facts about the world and the beliefs of cognitive agents who ask for explanations.

For example, to explain why Ms. R. voted for the left-of-center candidate in the latest election, one may cite the boundary condition that both her parents always did so, and the statistical law that 80 percent of voters vote for candidates from the same location on the political spectrum as their parents voted for. This form of explanation is thus an argument with two premises, one of which is a general law, or at least an empirical generalization that is well supported.

Explanans:

1. Eighty percent of voters vote for candidates from the same location on the political spectrum as the candidates that their parents voted for. (well-confirmed statistical generalization)
2. Ms. R.'s mother voted for left-of-center candidates. (boundary condition)
     Therefore,
        Explanandum:
3. Ms. R. voted for the left-of-center candidate in the latest election.

But clearly the argument form of this explanation is not deductive: The truth of the premises does not guarantee the truth of the conclusion: they are compatible with the woman in question not voting at all, or voting for the right-of-center candidate, etc.

Statistical explanations on this view are inductive rather than deductive arguments—that is, they give good grounds for their conclusions without

guaranteeing them. It is no defect of inductive arguments that they are not truth-preserving, do not provide guarantees for their conclusions (assuming the premises are true) the way deductive arguments do. All scientific reasoning from a finite body of evidence to general laws and theories is inductive—from the particular to the general, from the past to the future, from the immediate testimony of the senses to conclusions about the distant past, etc. (This is a matter on which we will focus in Chapters 10 and 11.)

In this case, the 80 percent frequency of voters voting as did their parents may be held to provide a 0.8 probability that Ms. R. can be expected to vote as she did. Thus, like D-N explanations, a so-called **inductive-statistical** (I-S) explanation gives good grounds that the explanandum phenomenon can be expected to occur. However, there is a serious complication that the I-S model must deal with. Suppose that in addition to knowing that both Ms. R.'s parents voted for candidates of the left, we also know that Ms. R. is a self-made millionaire. And suppose further that we know that it is a statistical generalization that 90 percent of millionaires vote for right-of-center candidates. If we know these further facts about Ms. R. and about voting patterns, we can no longer accept as an explanation of why she voted left that her parents did and that 80 percent of voters vote as their parents did. For we know that it is 0.9 probable that she voted for the right-of-center candidate. Evidently we need some other statistical or non-statistical generalization about female millionaires whose parents voted left to provide a statistical explanation for why Ms. R. did so. Suppose that the narrowest class of voters studied by political scientists includes female self-made millionaires from Minnesota, and that among these 75 percent vote for candidates of the left. Then we may be entitled to explain why Ms. R. so voted by inductively inferring from this generalization, and the fact that she is a self-made millionaire from Minnesota, that she voted as she did. This will count as an I-S explanation. Since this is the narrowest class of voters of which Ms. R. is a member, this is the one that explains her vote. So, to get an account of I-S explanation, we need to add to the four conditions on D-N explanation, something like the following additional condition:

5. The explanation must give a probability value for the conclusion no higher than the probability given in the narrowest relevant reference class the explanandum phenomenon is *believed* to fall into.

But notice that we have now surrendered a fundamental commitment of the logical positivists' approach to explanation: we have made the subjective beliefs of agents who ask for and offer explanations an essential element in scientific explanation, for it is our beliefs about the narrowest relevant reference class for which we have framed statistical regularities that determines whether an explanation satisfies the requirements of the I-S model. Of course, we could drop the qualification "is believed to" from (5), but if the underlying process that our statistical generalization reports is really a deterministic one, our I-S explanation will reduce to a D-N model, and we will have no account of statistical explanation at all.

The **counterexamples** that we have canvassed here led positivists and their successors to add clauses and conditions to their original rational reconstruction of scientific explanation. These additions were designed to preserve its features while accommodating those counterexamples that were held to reflect significant oversights in the positivist analysis of explanation. But in the aftermath of logical positivism, an entirely different approach to analyzing the nature of scientific explanation, and much else in science, emerged among philosophers of science.

A Competing Conception of Scientific Explanation

The logical positivists' commitment to rational reconstruction came with implicit prescriptive implications for how scientific explanation ought to proceed. This commitment persists among some philosophers of science long after the eclipse of logical positivism. We can contrast this commitment with a fundamentally different approach to the philosophy of science.

Some philosophers seek an objective relation between explanandum and explanans because they hold that science is constituted by truths about the world that obtain independently of our recognition, which we set out to uncover. Thus science is treated in the way that Plato and his followers down to the present conceive of mathematics, as the study of objective relations between abstract objects that obtain regardless of whether we recognize them. This approach to science may be more intuitively plausible than mathematical Platonism, if only because the entities that science seeks to uncover are not abstract (like numbers), but concrete (like genes).

By contrast with Platonism about mathematics, there are those who hold that mathematical truths are not about abstract entities and relations between them, but are made true by facts about concrete things in the universe, and reflect the uses to which we put mathematical expressions. Similarly, there are those who hold that science needs to be treated not like an abstract relation between truths, but as a human institution, a set of beliefs, and methods which we use to get around efficiently in the world. On this view scientific laws do not have a life of their own independent of the humans who invent and employ them. One might even try to capture this difference between philosophies of science by reflecting on the distinction between discovery and invention: Platonist-inclined philosophers treat the claims of science as truths to be discovered. By contrast there are the philosophers who treat science as a human institution, something we or the great scientists among us have invented to organize our experiences and enhance our technological control of nature. Platonists will seek an account of scientific explanation that makes it an objective relation between facts and/or statements that we set out to discover, while others seek a notion of explanation as an essentially human activity. The philosophy of science from which the logical positivist model of explanation emerges is one that treats science as an act of discovery, not invention. The contrasting approach treats science as a human activity, the result of our creativity and invention, perhaps even as a

construction. It takes seriously the epistemic limits and pragmatic interests of scientists and others who supply and demand scientific explanations.

The problems of statistical explanation and the flag-pole's shadow counter-example should lead us to take seriously such alternatives to the logical positivist theory of explanation, which emphasize the function of explanation in allaying human curiosity and conveying human understanding. Instead of starting with a strong philosophical theory and forcing scientific practice into its mold, these approaches are sometimes claimed to take more seriously what scientists and others actually seek and find satisfactory in explanations.

One way to see the differences between the pragmatic/epistemic approach to explanations versus the D-N approach is to consider that different explanatory requests can all be expressed in exactly the same words. A single syntactically and semantically identical expression can have different meanings. Suppose Ms. R. is caught red-handed standing over her husband's body with a revolver. Now consider the following three questions, spoken with different emphasis as indicated by the bold type.

(a) Why did **Ms. R.** kill Mr. R.?
(b) Why did Ms. R. **kill** Mr. R.?
(c) Why did Ms. R. kill **Mr. R.**?

To see the difference clearly, say each sentence out loud with the indicated emphasis.

Emphasis in speech is not a matter of syntax (grammar) or semantics (meaning), but of "pragmatics"—the way speech is used. The emphasis makes it clear that each question is a request for different information, and each presumably reflects differences in knowledge. Thus, the first presumes that Mr. R.'s being killed needs no explanation, only why it was Ms. R. instead of some other person "who did it" which needs explanation; the second question presupposes that what needs explanation is why what Ms. R. did to Mr. R. was a killing, and not a beating or a robbing, etc.; and the third question is a request for information that rules out other persons beside Mr. R. as the victim of Ms. R. Each of the different questions reflects one member of what Bas van Fraassen has called a "contrast class" of statements. Thus, the "contrast class" for (a) is {The *butler* killed Mr. R., the *cook* killed Mr. R., *Mr. R.'s daughter* killed Mr. R., *Ms. R.* killed Mr. R. …}. As expressed in (a), the request for explanation is in part a request to be shown why each of the other members of the contrast class can be excluded. The D-N model is blind to such differences in explanation that result from these differences in emphasis. As a result, some philosophers who reject logical empiricism advance an account of scientific explanation that starts with pragmatics.

Following an analysis of explanation due to Bas van Fraassen, call what the sentences (a), (b), and (c) above share in common the "topic" of the question. Now, we may associate with every question a three-membered set, whose first

member is its topic, whose second is the member of the contrast class picked out by the interests of whoever requests the explanation, and whose third member is a standard for what counts as an acceptable answer to the question, which is also fixed by the interests and information of the person seeking the explanation. Call this standard of acceptable answers to our explanatory question "the relevance relation," for it determines what answers will be judged as adequate in the context of the topic and the members of the contrast class in question. For an understanding of scientific explanation the crucial question is what sorts of things fill in the "relevance relation," what does science in general, or a particular science, or a particular scientist require to establish the explanatory relevance of any answer to the question expressing the explanatory inquiry.

Van Fraassen provides a partially symbolic formalization of this approach. We may identify every explanatory question with this set:

Q (why is it the case that Fab)? =	<Fab,	{Fab, Fac, Fad, …},	R>
	topic	contrast class	relevance relation

where "Fab" is to be read as a bears relation F to b; thus Fad means a bears relation F to d, etc. So if F is used to symbolize the property of "… is taller than …" then Fbc reads "b is taller than c." If F is used to symbolize the property of "… killed …" then Fab means a killed b, and so on. The question Q above is to be understood as including whatever emphasis or other pragmatic element is necessary to make clear exactly what is being asked. For example, "Why *Ms. R.* killed her husband" will be a different question from "Why Ms. R. *killed* her husband," and different from "Why Ms. R. killed *her husband*." All questions have (pragmatic) presuppositions ("Who let the dog escape again?" presupposes that the dog escaped and not for the first time, and that someone was responsible for allowing it). Explanatory questions are no exception. The presuppositions of Q include at least the following: that the topic, Fab (the description of what is to be explained), is true and that the other possibilities (the rest of the contrast class), Fac, Fad, etc. didn't occur.

Finally, the presuppositions of Q include the existence of an answer to Q, call it A. A explains Q if, in light of the background knowledge of the inquirer, there is some relationship between A and the topic, Fab, and the rest of the contrast class (Fac, Fad, etc.) that excludes or prevents the occurrence of the rest of the contrast class, and assures the occurrence of the topic, Fab. In our example, we seek a true statement which, given our knowledge, bears the *relationship* to the topic and the contrast class that it makes Ms. R.'s killing her husband true and the members of the contrast class false. As noted, van Fraassen calls this relationship between A and the topic and the contrast class "the relevance relation." We will want to know much more about this *relationship*. If our answer A is that

Ms. R. wanted to inherit Mr. R.'s money, then the background knowledge will include the usual assumptions about motive, means, and opportunity that are the police detective's stock in trade. If our background knowledge includes the fact that Ms. R. was rich in her own right, and indeed, much richer than her husband, the relevance relation will pick out another statement, for example, that Ms. R. was pathologically avaricious or terrifically jealous. Of course a scientific explanation will presuppose a different "relevance relation" than that involved in the crime reporter's explanation of why Ms. R. killed her husband. Van Fraassen tells us in effect that what makes an explanation scientific is that it employs a relevance relation fixed by the theories and experimental methods that scientists accept at the time the explanation is offered.

How does all of this apparatus enable us to improve on the D-N model? Because the analysis makes explanation openly pragmatic, it has no problem with the I-S model, nor with the notion that in different contexts explaining the flag-pole's height by appeal to its shadow's length will succeed. In the flag-pole example, if we know about the egalitarian and patriotic desires of the city mothers of Missoula, then the explanation in terms of the sun's rays, the size of the shadow, and the geometry of isosceles triangles will explain the height of the flag-pole. Similarly, in the I-S explanation, if we don't know that Ms. R. is a millionaire and/or we are acquainted with no further statistical generalizations about voting patterns, the initial I-S argument will be explanatory.

Independent of its ability to deal with any counterexamples, a pragmatic approach to explanation has its own motivation. For one thing, we might want to distinguish between a correct explanation and a good one. This is something that the D-N and I-S models cannot do, but which the pragmatic account can accommodate. Some true explanations are not good ones, and many good ones are not true. An example of the first kind frequently cited in philosophy explains to a child why a square peg will not fit in a round hole by appeal to the first principles of the quantum theory of matter instead of by appeal to facts the child is familiar with and can understand. An example of a good explanation though not a true one is provided by any of the well-confirmed but superseded theories that are part of the history of science. Physicists know well the defects in Newtonian mechanics. But Newtonian mechanics continues to provide explanations, and good ones at that.

The philosopher interested in *scientific* explanations will rightly complain that no matter what its other virtues, this pragmatic account does not illuminate scientific as opposed to other kinds of (non-scientific) explanations. In effect this pragmatic analysis of explanation leaves us no clearer than we were in the beginning on what makes an explanation scientific. All it tells us is that explanations are scientific if scientists offer and accept them. What we want to know are the standards for the "relevance relation" that will distinguish scientific explanations from the pseudo-explanations of astrology or for that matter the non-scientific explanations of history or everyday life. But if we cannot say a good deal more about the relevance relation, our analysis of explanation

will have little or no prescriptive bearing for how explanations ought to proceed in science, nor will it enable us to demarcate scientific from non-scientific explanations.

Summary

Our starting point for understanding scientific explanation is the deductive-nomological (D-N) or covering law model, advanced by the logical positivists. This analysis requires that scientific explanations satisfy the requirement of giving good grounds that their explanandum phenomena were to be expected. If we can deduce the occurrence of the event or process to be explained from one or more laws and boundary conditions, we will have satisfied this requirement.

Thus, the requirements for scientific explanation on this view are

1. The explanans logically implies the explanandum-statement.
2. The explanans contains at least one general law that is required for the validity of the deduction.
3. The explanans must be empirically testable.
4. The explanans must be true.

Each of these conditions raises serious philosophical problems.

One particularly important problem is that of exactly why laws explain. Laws are held to explain either because they report causal dependencies or alternatively because they express some sort of necessity in nature. This is the subject of Chapter 4.

Many explanations in physical science and most explanations in general fail explicitly to satisfy this model. Proponents of D-N explanation argue that explanations can in principle do so, and they should if they are to provide real explanations. Of course many explanations approximate to the D-N model, such as "explanation-sketches," are good enough.

Other philosophers reject both the D-N model and its motivation. Instead of a search for an objective standard against which to measure explanations for scientific adequacy, they focus on attempting to uncover the logic of the explanations that scientists—physical, biological, social, and behavioral—actually give. One reason to find this alternative strategy attractive arises when we consider the logical positivist account of statistical explanations, the inductive-statistical, I-S, model. Whether a statistical generalization is explanatory seems to be a matter of what is known about the population in the form of background information by those asking for the explanation and those offering it. This is difficult to reconcile with the D-N model.

Yet the alternative "pragmatic" approach to explanation does not successfully identify what distinguishes scientific explanations from non-scientific ones. This leads to problems about the laws and theories that confer explanation that we explore in the next chapters.

Study Questions

1. Defend or criticize: "The D-N or covering law model doesn't illuminate the nature of explanation. If someone wants to know why x happened under conditions y, it's not illuminating to be told that x is the sort of thing that always happens under conditions y."
2. Defend or criticize: "The D-N model represents an appropriate aspiration for scientific explanation. As such, the fact that it is not attainable is no objection to its relevance for understanding science."
3. Can the D-N model accommodate the flag-pole example?
4. Exactly where do the pragmatic and the D-N accounts of explanation conflict? Can they both be right?
5. Defend or criticize: "The pragmatic approach to explanation provides an account of how we use the words 'explanation' and 'understanding.' It does not illuminate what explanation and understanding really are."
6. What is the relationship between explanation, an epistemic activity, and causation, a process in the world?
7. Explanations are supposed to confer understanding. How, if at all, does this truism constrain the nature of scientific explanation?

Suggested Readings

Balashov and Rosenberg's *Philosophy of Science: Contemporary Readings*, the anthology designed as a companion to this text, includes several of the important papers on explanation that have influenced discussion of these topics over the last 50 years. See section II, "Explanation, Causation and Laws." Some of these papers by Hempel and others are also to be found in two other anthologies: Marc Lange, *Philosophy of Science: An Anthology*, and M. Curd and J. A. Cover, *Philosophy of Science: The Central Issues*. The latter volume provides especially cogent editorial essays explaining and linking the articles.

The debate about the nature of explanation begins with classic papers written by Carl G. Hempel in the 1940s and 1950s and collected together with his later thoughts in *Aspects of Scientific Explanation*. Much of the subsequent literature of the philosophy of science can be organized around the problems that Hempel raises for his own account and deals with in these essays. The final essay in Hempel's book addresses the work of other philosophers who responded to Hempel's account. Balashov and Rosenberg reprinted Hempel's paper outlining the D-N and the inductive-statistical accounts, "Two Models of Scientific Explanation." Hempel's "Inductive-Statistical Explanation" is reprinted in Curd and Cover.

The subsequent history of debate about the nature of explanation is traced in Wesley C. Salmon, *Four Decades of Scientific Explanation*, originally published as a long essay in volume 13 of *Scientific Explanation*, in the Minnesota Studies

in the Philosophy of Science, W. Salmon and P. Kitcher, eds., and subsequently published as a separate volume. The volume from which it comes is a treasure trove of contemporary papers on the nature of scientific explanation.

Van Fraassen's approach to explanation is developed in *The Scientific Image*, from which an extract is provided, "The Pragmatics of Explanation," in Balashov and Rosenberg. P. Achinstein, *The Nature of Explanation*, advances a pragmatic theory of explanation that differs from van Fraassen's.

J. Pitt, *Theories of Explanation*, reprints many important papers on explanation, including Hempel's original paper, Salmon, "Statistical Explanation and Causality," P. Railton, "A Deductive-Nomological Model of Probabilistic Explanation," van Fraassen, "The Pragmatic Theory of Explanation," and P. Achinstein, "The Illocutionary Theory of Explanation."

The most influential recent work on explanation is Jim Woodward, *Making Things Happen*, which carries on Salmon's causal approach. L. Paul and N. Hall, *Causation: A User's Guide*, expands on Woodward's approach while exploring its metaphysical foundations. Marc Lange, *Because without Cause: Noncausal Explanations in Science and Mathematics*, explores important alternative approaches to explanation.

Other important papers on explanation are mentioned in the reading guide at the end of the next chapter, also devoted to explanation.

4   Why Do Laws Explain?

Overview

Regardless of which approach we adopt to the nature of scientific explanation, we still need to address the question of why laws explain. Undoubtedly some scientific explanations can do without explicitly citing laws, and sometimes regularities that are not really laws of nature are enough to provide some scientific understanding. But even when not stated, laws always seem to be somewhere in the background making the explanatory connections. So we still need to understand why laws have a particularly central place in science. What is it about laws of nature that makes them explanatory? Why does scientific understanding centrally involve the search for laws of nature?

In this chapter we locate what has widely been agreed to be the source of laws' powers to explain: their necessity. This chapter introduces the several ways that philosophers have tried to deal with the grave metaphysical problems about science and reality that are raised by the power of laws to explain. These metaphysical problems arise repeatedly in the philosophy of science, due to the centrality of laws to the enterprise of science.

What Is a Law of Nature?

The logical empiricists identified several features of a law on which there has continued to be widespread agreement: laws are universal statements of the form "All As are Bs," or "Whenever an event of type C occurs, an event of type E occurs," or again, "If (an) event e happens, then invariably, (an) event f occurs." For example, "All pure samples of iron are conductors of electric currents at standard temperature and pressure," or "Whenever an electric current is applied to a sample of iron under conditions of standard temperature and pressure, the iron conducts current," or "If an electric current is applied to a sample of iron under standard temperature and pressure, then the sample conducts current." These are terminological variants of the same law. Philosophers tend to prefer the "if …, then …" conditional version to express the form of a law, making the universality of the law implicitly understood. Merely being a statement of universal form, however, is not sufficient for being a law. Consider "All bachelors are unmarried," and "All even numbers are divisible without remainder by 2." Neither of these universal statements is a law, even though both have the same logical form as laws. They are not laws because they are true by definition, whereas laws are made true by facts about the world.

Besides their form, another feature that all laws share is that they don't refer to particular objects, places or times, either implicitly or explicitly. There can't be laws of nature that are made true just by the facts about Napoleon Bonaparte or the Earth or even the Milky Way. Laws are supposed to hold everywhere and always, so no local facts could be enough to make them true. Indeed, there are laws of nature that are true even though no objects in the universe make them true. Consider three laws about Bohrium, a synthetic element: It has a half-life of 61 seconds, has an atomic weight of 270, and each of its atoms has two outer-shell electrons. But it is quite probable that there is at present not a single atom of Bohrium anywhere in the universe. It has to be synthesized in a particle accelerator. But the laws about it, and the other synthetic elements, "obtain" and are true in spite of the fact that there are no examples of its apparent subject.

These features—logical form, making contingent claims about nature, and freedom from local facts—are not sufficient to distinguish laws from other statements that are grammatically similar to laws but lack their explanatory force. The real difference between laws and other statements that look like them is therefore obvious, yet hard to understand. Laws explain because laws have some sort of *necessity to them*. There is some sense in which they tell us how things *must be* arranged, how they *have to be* organized, not just how as a matter of fact they happen to be arranged or organized. The trouble comes in figuring out exactly what kind of necessity it is that laws have. It cannot be logical necessity, as dictated by the laws of logic, or the necessity of definitions. Laws, as we saw in Chapter 2, are not guaranteed by logic. The universe could have perfectly well been governed by an inverse cube law of gravitation instead of an inverse square law, without violating the laws of logic. The trouble with saying that laws are necessary, as we shall see, is to figure out exactly what sort of necessity it is

that laws have. Laws need necessity in order to be explanatory, but what sort of necessity might this be?

Counterfactual Support as a Symptom of the Necessity of Laws

To see why laws have some sort of necessity, compare the following two statements, each of which has the same universal form:

> All solid masses of pure plutonium weigh less than 100,000 kilograms.
> All solid masses of pure gold weigh less than 100,000 kilograms.

We have good reason to believe that the first statement is true because quantities of plutonium spontaneously explode long before they reach this mass. Thermonuclear warheads rely on precisely this fact. There is also good reason to think that the second statement is true, but it is true just as a matter of cosmic coincidence. There could have been such a quantity of gold so configured somewhere in the universe, but there is not. Presumably the former statement reports a natural law, while the latter describes a mere fact about the universe that might have been otherwise. One way to see that the statement about plutonium is a law is that an explanation of why it is true requires us to appeal to several other laws but no initial or boundary conditions; by contrast, to explain why there are no solid gold masses of 100,000 kilograms requires laws and a statement of boundary or initial conditions that describe the distribution of atoms of gold in the universe from which gold masses are formed. What this shows is that universality of form is not enough to make a statement a law of nature.

One symptom of the difference between real laws and accidental generalizations that philosophers have hit upon involves grammatical constructions known as "**counterfactual conditionals**," or "counterfactuals" for short. A counterfactual is another sort of if/then statement, expressed in the subjunctive tense. It concerns what *might have been* the case if something else (that happens not to be the case) had instead obtained. We employ such statements often in everyday life: "If I had known you were coming, I would have baked a cake." Two examples of such counterfactual statements relevant for distinguishing laws from non-laws of the same grammatical—"if …, then …"—form are the following:

> If it were the case that the Moon is made of pure plutonium, then it would be the case that it weighs less than 100,000 kilos.
> If it were the case that the Moon is made of pure gold, then it would be the case that it weighs less than 100,000 kilos.

Notice that the antecedents (the sentences following the "if"s) and the consequents (the sentences following the "then"s) of both counterfactuals are false. This grammatical feature of counterfactual sentences is obscured when we express them more colloquially and less stiltedly as follows:

If the Moon had been composed of pure plutonium, it would weigh less
than 100,000 kilos.

If the Moon had been composed of pure gold, it would weigh less than
100,000 kilos.

So, these two statements are claims not about actualities but possibilities—the
possible states of affairs that the Moon is composed of plutonium and gold
respectively. Each says that if the antecedent obtained (which it doesn't), the
consequent would have obtained (even though, as a matter of fact, neither does
actually obtain). Now, we hold that the counterfactual about gold is false, but
that the counterfactual about plutonium is true. And the reason for this diffe-
rence is that the universal truth about plutonium supports the plutonium coun-
terfactual, while the universal truth about gold masses does not support the
gold counterfactual. Why not? Because it is not a law, but merely an accidental
generalization.

Thus, we may add to our conditions on laws that in addition to being universal
in form, they must support their counterfactuals. Yet it is crucial to bear in mind
that this is a *symptom* of their being laws, not an explanation of it. That is, we can
tell the difference between those generalizations that we treat as laws and those that
we do not by considering which counterfactuals we accept, and which we do not
accept. But unless we understand what makes counterfactuals true independent of
the laws that support them, the fact that laws support their counterfactuals won't
help to explain the difference between laws and accidental generalizations.

We know that laws support their counterfactuals, while accidental
generalizations do not. Presumably, laws support their counterfactuals because
laws express some real, necessary connection, some glue, that holds instances of
their antecedents and their consequents together that is missing between the ante-
cedent and the consequent of an accidental generalization. Thus, there is some-
thing about being a sphere of pure plutonium that brings it about, or necessitates,
the fact that it cannot be 100,000 kilos in mass, whereas there is nothing about
being a sphere of gold that makes it impossible for it to be that massive.

But what could this glue be—this real connection between the antecedent
and the consequent of a law, which reflects the necessitation of the latter by the
former? As the example of the inverse square law of gravitation shows, laws do
not express logical necessities. Or at least this is widely believed in the philosophy
of science, on the grounds that denial of a natural law is not itself contradictory,
whereas the denial of a logically necessary statement, like "All even numbers are
divisible without remainder by 2," *is* contradictory. It's impossible to conceive of
the violation of a logically necessary truth. It is easy, however, to conceive of the
violation of a natural law. Laws of nature cannot be logically necessary.

Well then, what if we said that what distinguishes scientific laws from other
statements of the same grammatical form that don't support counterfactuals is
that laws are physically or perhaps chemically or biologically necessary? More
broadly that they are "naturally necessary"? Violations of the laws of nature
are not logically impossible, but only physically impossible. The idea is that

physical impossibility is like logical impossibility, only weaker. Logical impossibility is the strongest sort of impossibility—a logical impossibility cannot even be imagined coherently because it is ruled out by the laws of logic. For example, an odd number divisible without remainder by 2. But a physical impossibility is weaker: a physical impossibility is ruled out by the laws of physics. A chemical impossibility is anything incompatible with the laws of chemistry. A biological impossibility is what the laws of biology forbid. For example, traveling faster than light, a molecular bond between two rare earth elements, or the persistence of a dominant lethal gene. Usually the term "physical impossibility" is employed to describe incompatibility with any natural law—physical, chemical, biological, etc. Of course, there will be even weaker impossibilities, technological or practical impossibilities, ruled out not by the laws of nature but by the limits of our technology: for example, tabletop nuclear fusion. But this nice grading of different strengths of possibility and impossibility won't really help us to understand natural necessity. In fact it enables us to see the problem with special clarity.

It's no explanation of the necessity of laws to say they reflect "physical" or "natural" or even better "nomological" (from *nomos*, Greek for law) instead of logical necessity. What is it for a statement to be a physical or natural necessity except that it is required to be the case by the laws of physics or nature? If this is what natural or physical necessity consists in, then grounding the necessity of laws on physical or natural necessity, is grounding the necessity of laws on itself!

Counterfactuals and Causation

This question of what kind of necessity laws have, and accidental generalizations lack, is exactly the sort of "metaphysical" question that the logical positivists hoped to avoid in their analysis of explanation. For nomological necessity just turns out to be the same thing as the necessity that connects causes and their effects and is missing in merely accidental sequences. Thus logical positivists were unable to avoid dealing with the metaphysical problem raised by causality. We have no aversion to metaphysics so perhaps we can make progress in understanding what makes a generalization a law by thinking more about causality. At a minimum the sameness of the necessity of laws and the necessity of causation will illuminate what it is about covering law explanations and causal explanations that make them explanatory.

Recall our brief discussion in Chapter 3 of causal sequences versus coincidences. Presumably a causal sequence is one in which the effect is brought about by the cause, produced by it, made to happen by the cause's occurrence, necessitated by it. One way of making this point is to put it this way: "If the cause hadn't happened, the effect would not have happened"—the counterfactual kind of statement we encountered when trying to understand the necessity of laws. By contrast to a causal sequence, there is no such relation of necessitation between the first event and the second in a coincidental sequence. But what does this causal necessitation consist in? As David Hume suggests, there does not seem to be any "glue" or other observationally or theoretically detectable connection

between causes and their effects anywhere in the universe. All we ever see, even at the level of the microphysical, is one event followed by another.

Try the following thought experiment: consider what goes on when one billiard ball hits another and the second one moves. The transfer of momentum from the first to the second is just a way of saying that the first one moved, and then the second one did. After all momentum is just (mass × velocity) and the masses didn't change, so the velocity must have changed when the momentum was transferred. Consider the counterfactual that "If the momentum hadn't been transferred to the second ball, it would not have moved." Why not? Will it help to consider what happened at the level of the molecules out of which the billiard balls are made? Well, the distance between them became smaller and smaller until suddenly it started to grow again as the balls separated. But there was nothing else that happened below the level of observation besides the slowing down of forward motion of the molecules in the first billiard ball, followed by the speeding up of the forward motion of the molecules that made up the second. Nothing, so to speak, hopped off of the first set of molecules and landed on the second set; the first set of molecules didn't have a set of hands that reached out and pushed the second set of molecules. But if we try the thought experiment at a deeper level, say the level of atoms, or the quarks and electrons that make up the atoms, we will still see only a sequence of events, one following the other, only this time the events are sub-atomic. In fact, the outer-shell electrons of the molecules on the surface of the first ball don't even make contact with the electrons on the outer shells of the molecules at the nearest surface of the second ball. They come close and then "repel" each other, that is, move apart with increasing acceleration. No matter how far down we go in the details, there does not appear to be any glue or cement that holds causes and effects together that we can detect or even imagine.

If we cannot observe, detect, or even conceive of what the necessary connection between individual instances of causes and their effects might be, the prospect for giving an account of how causal explanation works or why laws have explanatory force becomes dimmer. Or at least the logical positivists' hope to do this in a way that avoids metaphysics will be hard to fulfill. The difference between explanatory laws and accidental generalizations, and the difference between causal sequences and mere coincidences, appears to be some sort of necessity that the sciences themselves cannot uncover. If the question of why laws explain has been answered by the claim that they are causally or physically or nomologically necessary, the question of what causal or physical or nomological necessity most basically *is* remains as yet unanswered. Answering this question takes the philosophy of science into the furthest reaches of metaphysics, and epistemology, where the correct answer may lie.

Coming to Grips with Nomic Necessity

Logical positivists may have refused to discuss the nature of nomological (or synonymously "nomic") and causal necessity on the grounds that metaphysics lacked cognitive significance, but once we accept that laws really do support

counterfactual conditionals, while accidental regularities don't, we cannot avoid the challenge of providing an account of what this difference comes to.

The earliest attempts to do so, by former positivists and empiricist philosophers (A. J. Ayer chief among them), tried to substitute for the task of identifying some "metaphysical" fact about natural laws a different task: find some epistemic difference in our attitudes or beliefs about the laws themselves. Their idea was simply to treat the symptoms of nomic necessity as the very thing that nomic necessity consists in. The difference between laws and accidental regularities was held to be a matter of how they were treated by scientists: if a regularity is accepted after only a small number of experiments, if it is accepted as explaining the data in those experiments, if it supports counterfactuals about events and processes not observed, and if it can be explained by more general regularities, then it is a law. Otherwise, it is merely an accidental regularity.

The trouble with this approach is that it mistakes the laws of nature for our best scientific hypotheses about what the laws of nature are. Of course we treat our best guesses about the laws of nature in the way described above: we accept them after only a few replications of well-designed experiments, we use them in explanations, try to link them to other hypotheses we accept, and assign truth or falsity to counterfactual conditions in accordance with them. But science is fallible and we often turn out to be wrong in identifying the laws of nature. But mustn't a law of nature be true? We noticed this problem in connection with the covering law model's demand that in a scientific explanation the explanans, including the law, be true. The way we treat our best guesses can't be what distinguishes laws that we might or might not even know yet from accidental regularities.

The problem is to find some property or feature of laws—known or unknown— that distinguishes them from mere regularities. An epistemological solution to a metaphysical problem is rarely satisfactory, as we shall have occasion to see hereafter.

Many philosophers recognized that the problem of nomic necessity required a metaphysical solution but were reluctant to engage in what the positivists might have stigmatized as extravagant speculation that was lacking in empirical content. They did not want to add to the list of necessities besides logical necessity, which seems to be well understood. Those philosophers sympathetic to empiricism advanced an account of what makes something a law of nature that required nothing more than logical necessity and the existence of propositions or statements about reality that exist independent of anyone who might take notice of or believe them. Notice, propositions are not sentences. Sentences in a particular language express propositions, but are not identical to them. Consider, "It is raining," "Es regnet" and "Il pleut." These are three sentences that all express the same proposition, which presumably is distinct from any of the three sentences. The existence of propositions distinct from sentences seems to be a relatively uncontroversial claim once the existence of abstract objects of any kind is admitted. But it is not wholly without controversy, since once one sort of abstract object is accepted in an "**ontology**"—a list of the kinds of things

there are in reality—other such objects may have to be accepted: properties independent of things, so-called "universals," numbers as distinct from numerals and concepts of numbers, etc.

How can the existence of propositions help distinguish laws from merely accidental regularities? Consider all the local matters of particular fact from the beginning of time, or from the Big Bang onward, that constitute the history of the universe. This set of facts may or may not obtain over an infinite length of time and may or may not include an infinite number of events. Regardless, there will be many sets of propositions that describe some or all of these facts. Some sets will describe them all, but be completely unsystematic in their relationship to one another, rather like a list of unrelated facts in an almanac. These sets of statements are strong on content but lack any systematic simplicity or economy in their description of the universe.

Other sets of propositions will not be so complete in their descriptions of the local matters of particular facts. But they will be organized into logical systems that express relations of implication between statements, rather like the axioms and theorems of Euclid's *Elements*. From the axioms all the other propositions in the set will be logically derivable. Such a system will not describe all the local matters of particular facts that obtain in the universe's history, but it will be much more economical and systematic than the complete "almanac description."

There will be many such sets of axioms that describe more or less the sum-total of local matters of particular fact. The more local facts that one of these deductive systems of axioms and theorems describes, the more complicated and numerous will its axioms have to be and the fewer theorems it will need to describe the facts; the fewer and simpler the axioms, the more theorems it will need and the more local facts it will leave out of its description. Now, consider those axiomatic systems that optimally combine simplicity and strength in their descriptions of the local matters of particular facts. Among these "best" axiomatic systems there will be many statements of universal form, "All As are Bs" or "Whenever *c* then *e*" or "If F then G." The laws of nature will be those propositions that are axioms or theorems of *all* of these systems of statements that optimally combine simplicity and strength in the description of local matters of particular fact. Presumably the general statement about plutonium spheres above will be among the theorems of every axiom system tied for simplicity and strength (because it follows from quantum mechanics). By contrast the general statement about gold spheres above will not be among the axioms or the theorems of any of these best systems. If we want to include it in any axiomatic system, we will have to add it as an axiom, and doing so will greatly reduce simplicity without much gain in description of local matters of particular fact.

So, a sentence in any language expresses a law if it expresses a proposition of universal form that is either an axiom or a theorem in all of the deductive systems of propositions that are tied for being the best combination of simplicity and strength in describing all the local matters of particular fact in the history of the universe. This analysis of what nomological necessity consists in is often called the "best-systems" account. Its proponents go back from the

contemporary philosopher David Lewis through the post-World War I British philosopher Frank Ramsey, back to the nineteenth-century empiricist John Stuart Mill.

This account of nomological necessity has many attractions. To begin with, it connects being a law with larger units of scientific description, since the axiomatic systems every law belongs to are in fact "theories of the world," systems that describe and explain (if explanation is deduction) all regularities among local matters of particular fact in the history of the universe. Second, this approach analyzes away the notion of natural or nomic physical necessity into the well-understood notion of logical necessity by appealing only to axiomatic systems in which statements are related to one another by logically necessary deductive relationships. Third, it is metaphysically modest: it does not hypothesize some empirically undetectable glue holding together the lawfully connected events, but absent from the merely coincidental sequences of events.

The best-systems theory of nomic necessity faces a number of significant challenges. What these objections have in common is largely the criticism that the best-systems view is not sufficiently different from the post-positivist view of A. J. Ayer and other empiricists like him. The best-systems view (like Ayer's view), it is argued, is not sufficiently "mind-independent." It does not provide the identity of physical necessity in some metaphysical facts "in the objects," in the local matters of particular facts, or in any other facts about the world independent of our beliefs about it. It may seem to do so, since statements and propositions along with the logical relationships among them obtain regardless of whether any cognitive agent thinks of them and their logical relations. But once simplicity and strength are introduced as a set of standards for choosing the "best systems of propositions" we have to introduce sentences of languages, and neither sentences nor the languages that express them are mind-independent in the required way.

Why do we need to measure simplicity and strength as properties of language? Consider simplicity. Can we characterize this feature of axiomatic systems without thinking about the purposes cognitive agents put them to? One system is simpler than another presumably if it contains fewer axioms. But any system of propositions, no matter how many it contains, can be turned into a system with just one axiom—the conjunction of all its propositions, or the conjunction of all the axioms that logically entail the rest of its propositions. Such axiomatic systems are not the simplest ones in the sense required by the best-systems theory. The best system is the one that optimally combines simplicity and strength for the uses to which a cognitive agent might put the system in explaining as many of the local matters of particular fact as possible. If explanation is fundamentally a pragmatic and epistemic enterprise, the best-systems approach turns out to be not much better than the empiricist approach that changes the subject from laws to hypotheses, then tells us the difference between our hypotheses and merely accidental regularities. And if explanation is by covering law, then the best-systems analysis is going to require that we state propositions in particular languages and show that deductions—of explananda

from explanans—are valid. Either way, simplicity as a feature of systems is going to be a matter of the judgment of scientists, instead of facts about the world independent of their thoughts.

There are more obvious objections to the best-systems approach: What if there is no objective standard of simplicity and/or strength? How could we settle a disagreement between a Kantian (who would hold that simplicity is more important than strength) and an empiricist (who would hold that strength is more important)? Simplicity and strength are notoriously vague terms. If their vagueness cannot be eliminated, will it turn out that being a law of nature is a vague category with borderline cases, the way that being a short person is a vague category with borderline cases? There does not seem to be any other reason to suppose the category of being a law of nature is vague, so this is a serious matter. Worse, what if it turns out that several systems of axioms are tied for optimal simplicity and strength in describing local matters of particular fact, and they have no or few propositions in common as axioms and theorems? Are we to say that there are no laws of nature or only a few laws?

These objections will begin to sound less serious when we consider an alternative account of nomic necessity, one with a pedigree in philosophy that goes back to Plato! This alternative is on the one hand metaphysically extravagant in a way that few contemporary philosophers of science could be comfortable with, but at the same time rests on some important features of laws of nature that no other theory seems to be able to account for at all.

"All and only mammals bear their young alive" seems to be a law, as does "All mammals have four-chambered hearts." From these two apparent laws, it follows logically that "All animals that bear their young alive have four-chambered hearts." But this latter may not seem to you to be a law. It is true of course, and true in virtue of the truth of two laws. But there doesn't seem to be anything about the property of being mammalian that links these properties together in the way a law requires. How did we get from two laws to a non-law by a valid argument? All we did was substitute for the property of being a mammal the property of bearing one's young alive. These two properties are shared by exactly the same set of things, the mammals. And yet that changed a law into a non-law. If laws are about the particular things in the world they describe, this shouldn't happen, should it? What if laws are not really about things in the world—local matters of particular fact? What if they are really about *properties*? This idea is the core of a metaphysical account of nomic necessity that goes back to Plato, 2,400 years ago.

Here unfortunately we need to introduce some terminology. Plato and many philosophers since he wrote have argued that there are properties and they are not to be confused with the property-word (predicates) we use to name them. The distinction is the same as that between sentences in a language and the propositions they express. "Red," "rouge," "rot," "rojo" are words: predicates that all describe the same property. That property is something shared in common by all red things. Plato held that the property "redness" named by the predicates and shared by red things is itself another thing, an abstract thing (along with

numbers and propositions). To emphasize the contrast between, for example, particular red things and redness, philosophers refer to properties like redness as "universals." The thesis that there are universals, and that they (as well as numbers and propositions) have an existence independent of particular things that exemplify them (or in the case of numbers are counted by them and in the case of propositions make them true), has for centuries been called "realism," short for "Platonic realism." (This nomenclature is somewhat confusing, since many philosophers have denied the existence of abstract objects out of what they insist is a "robust sense of realism." Moreover, in Chapter 8 we shall encounter another use of the term "realism" in "scientific realism"—the label for a thesis in quite a different controversy in the philosophy of science.) Platonic realism, however, is widely recognized among philosophers to be the doctrine that properties, numbers, and propositions have an existence independent of their instances, the things they count, and what makes them true. Many philosophers not otherwise followers of Plato have accepted this thesis over the years. Empiricists almost all reject the thesis. They are "nominalists," holding that predicates—the property-words in language—don't pick out, refer to, or require the existence of real abstract objects—properties, numbers, propositions independent of concrete things and facts.

How can realism about universals help us to understand nomic necessity? Suppose, as some influential philosophers of science argue, that laws don't describe relations between things in the world, local matters of particular fact. Instead, suppose laws describe relations between universals, global matters of abstract fact. Thus, "Water expands when frozen," states that the properties of being water and of expanding when frozen "go together," above and beyond the fact that their instances go together. On this view, it's because the abstract properties "go together," that their instances—particular cases of being water and of expanding when frozen—*must go* together. The fact that the universals in the law "go together" confers necessity on all particular instances of "being water" co-occurring with particular instances of "expanding when frozen."

This theory has several things to be said for it. First, it will be no surprise that we do not find any "glue" among the local matters of particular fact that are connected by laws, since the real work of conferring nomic necessity is accomplished by the relation between the universals, not their particulars. Second, it will be obvious why switching properties that are different but have exactly the same instances can change a law into a non-law. The theory also will have no problem explaining why laws support counterfactuals and accidental regularities do not. For example, the theory will explain the truth of the counterfactual that plutonium masses of a certain size would explode, while similar counterfactuals about being a gold mass of a certain size and exploding are not true. It will inform us that the properties, the universals, of being a plutonium mass of a certain size, and of exploding, are related to one another by an equally abstract relationship of "going together" or perhaps "being causally related," or "bound together in nomic necessitation."

Another advantage of realism is that universals, like numbers and propositions, if they exist at all, exist quite independent of the thoughts of any cognitive agents. So this approach avoids the problems of the best-systems theory and its empiricist predecessor. Laws are about independently existing universals, and they are "out there" waiting to be discovered by inferences from observation of their particular instances.

Realism's claim that laws are relationships between universals is obviously a metaphysically extravagant theory. It faces many questions owing to this extravagance. Perhaps the first question it faces is the one that Plato faced when he sought to explain how the universal "being red" shared by all red things in common could have anything to do with their being red! After all, the universal "redness" is abstract, it is not in space and time, it can have no causes or effects, it is causally "inert." For this reason Aristotle rejected Plato's idea that there are such abstract universal properties at all.

The same problem and several others like it are faced by the contemporary appeal to universals. The theory needs to explain how any relationship of one universal to another translates into the relationship between their particular instances in the real world and makes that relationship different from the relationship between instances of properties that are merely accidentally related to one another. Silence on these matters leads opponents of realism to complain that this theory just adds mysteries to the problem of distinguishing laws and accidental regularities.

And there is a second problem. Assuming that there are universals, exactly what is the relationship between them, labeled here "going together," that makes the statement that they go together a law? "Goes together" is itself a metaphor, since on its literal meaning, things go together only if they are co-located—occur at the same time and place. But universals are abstract. They do not exist at any time or place, or alternatively they all exist at every time and place. So, on its literal interpretation, "goes together" doesn't really illuminate how universals do their work of conferring necessity on the co-occurrence of their particular instances.

Some proponents of this theory about laws claim that in a law it is the universals that are necessarily connected, and that is what confers necessary connection on their particulars. This view is not popular since it simply raises the question of what sort of necessity laws have all over again. Since we can conceive of the falsity of any law of nature, it is often held that laws cannot be logically necessary. Suppose, however, that conceivability is not a guide to logical possibility so that conceiving of, say, an inverse cube law of gravitational attraction does not show its logical possibility. In that case, it might still turn out that laws of nature are necessary in the same way that the laws of logic are. Now of course this leads to a new problem: figuring out what makes the laws of logic and of science necessary truths, since we can no longer use inconceivability as a test for violations of logical or any other kind of necessity.

Nevertheless, in recent decades some philosophers have argued that at least some laws must share the sort of necessity that the laws of logic have. The

restricted set of laws they hold to be necessarily true with the same strength and sense of necessity as the laws of logic are ones that report structural identities such as "All samples of water are samples of H_2O" or "Heat is identical to molecular motion." The argument for this claim is not scientific. It turns on a theory about how words like "water" and "heat" refer to or name their instances. Following three philosophers who developed this theory separately, Ruth Marcus, Hilary Putnam, and Saul Kripke, it is argued that "water" is defined by reference to some particular sample of the stuff we run into on Earth. Since the sample we or whoever coined the term "water" ran into is in fact H_2O we won't count anything anywhere else in the universe or even in other possible (conceivable, non-actual) worlds as water unless it has the same structure, is composed of H_2O. Accordingly, the law that water is H_2O must be a necessary truth because it is true in every possible world, just like the laws of logic. This theory accepts that learning that water is H_2O requires a good deal of experiment and observation. Unlike the laws of logic, these laws reporting necessities of identity are not known *a priori*, but they are still just as necessarily true.

Can this theory, that some laws reporting structural identities are (logically or "metaphysically") necessary, even though their denials are conceivable, help clarify the nature of the connection between universals or any other account of nomological necessity? Since the theory applies only to laws that record identities, its details are of limited immediate application. But in fact, it appears to make the entire problem of the nature of necessity much more difficult. We thought we had a handle on logical necessity as reflecting the operation of the laws of logic. They are the laws we know *a priori* through attempts to conceive of their denials. If we can no longer treat conceivability as a test of logical or metaphysical possibility, and inconceivability as a test of logical or metaphysical necessity, and have to assimilate the laws of logic to the laws of nature, our problem has become more serious, not less.

Denying the Obvious?

The problems raised by the distinction between laws and accidental regularities are daunting. They will certainly enable us to sympathize with and appreciate the logical positivists' attempt to exclude such questions as mere speculative metaphysics in the pejorative sense of that term. Unfortunately, once we give up any principle of verification, we have no grounds to exclude such questions. But perhaps we can do without the distinction altogether. This at any rate has been the view of at least some philosophers of science.

Some of these philosophers will want to deny the existence of abstract objects or at least their role in any kind of science or even in thought. They will deny that we need to accept propositions, numbers, or properties in our metaphysics or that understanding science requires any appeal to metaphysics. For these philosophers, things like sentences, numerals, and predicates are enough. They will hold that the difference between laws and accidental generalizations is just given by the differences between the way we treat some sentences and not others.

Sentences we accept after little experimentation, that we appeal to in order to ground counterfactuals we think true, that satisfy our explanatory curiosity, may be counted as laws. Of course we and science are both fallible; over time we both change our minds about which hypotheses we treat as laws. But being treated as a law seems, to these philosophers, to be all that a regularity needs in order to do the work that laws do in science. These philosophers claim inspiration from Hume, who at some places in his writings held that there was nothing more to being a law than being a true statement of constant conjunction between events or states of affairs. If, as a matter of fact, every gold sphere that has ever existed or will exist in the whole history of the universe is smaller than 1 kilometer in diameter, then this is a law of nature. Its failure to support counterfactuals simply shows that we retain an illusion that there is some necessity in the objects. Philosophers and scientists adopting this view will have neatly disposed of the problem of accounting for nomic necessity simply by denying that they must do so. But now they have no resource for illuminating the indispensability of laws in scientific explanation.

By contrast with those who seek to preserve the role of laws while rejecting nomic necessity, some others have sought to reject laws but hang on to necessity. These philosophers dispose of the problem of distinguishing laws from accidents by denying that there are laws in nature at all. One prominent advocate of this view has been Nancy Cartwright. Cartwright is prepared to deny that scientific explanations invoke any exceptionless universal regularities that obtain in nature. This includes Newton's laws, or their quantum mechanical relativistic successors. Of course we employ the sentences that seem to express these laws in scientifically satisfactory explanations. But the explanations provide no reason to think the "laws" they mention are true about the world. Take for example the inverse square law of gravitational attraction. Since everybody in the universe is also subject to electrostatic laws, explanations of their behavior that appeal only to gravity will be incomplete. Predictions of the behavior of bodies will be wrong for the same reason—perhaps only very slightly, and inconsequentially for explanatory purposes. But, Cartwright argues, the real force always acting on bodies is both gravitational and electrostatic, and the notion that there is a distinct solely gravitational law ever realized by local matters of particular fact is just false. In scientific practice we abstract from the force in operation, calculate first the gravitational effects, then the electrostatic effects, and combine them if they are strong enough to both have a significant effect on a body's path. But, Cartwright insists, nature doesn't work that way. It does not consist in two separate kinds of forces that always operate together; these are scientists' abstractions (not to be confused with abstract objects).

Instead of a law of gravity operating on everybody in the universe, everybody in the universe has a capacity or a **disposition** to exert gravitational forces and be subject to them. These dispositions are more fundamental than the regularities that they give rise to, which, if true, are much more complicated than any law we might find in an actual scientific explanation.

What exactly is a disposition? A disposition is something that an object may have even when it is not manifesting it. A magnet has the disposition to attract iron filings—it's magnetic, even when there are no iron filings in its vicinity to attract. An inflammable liquid has the disposition to burst into flame when a spark is applied to it, but may never burn, because no such spark was applied. Sugar is soluble, but some cubes of it may never be placed in a liquid and so never dissolve. Many important properties that science has discovered are dispositions. Notice that if something has a disposition then saying so supports counterfactuals about it—a dead give-away that dispositions harbor the sort of natural necessity we have been hoping to illuminate. A simple illustration reveals this. If a piece of iron is magnetic (a disposition), then it would produce an electric field if moved through a coil. Why is this counterfactual true? Some philosophers will say because there is a law that grounds magnetism and electricity. Cartwright and other philosophers argue that matters are the other way around: it is the dispositions that are more basic, that are present in objects and that we can detect in them, that bear the relevant sort of necessity, and that ground our incomplete but useful scientific laws.

Traditionally, empiricist philosophers have been wary of dispositions: the claim that a thing has a disposition but does not manifest or express it often or ever seems to be difficult to test and therefore easy to assert without providing much or any evidence. Empiricists often make fun of explanations of actual events that appeal to dispositions. They quote the French playwright Molière, who ridiculed a character for explaining why opium makes people sleep: "Opium has a dormative virtue—a disposition to make people sleep." Unless we can locate in samples of opium a non-disposition, a separate detectable trait opium manifests all the time—for example its chemical structure—that its disposition "rests on," the dispositional explanation is threatened with vacuity.

For this reason empiricists have argued that scientifically acceptable dispositions must be understood in terms of laws about the make-up of the objects that have these dispositions, and about empirically testable laws about these components or constituents of the things that give rise to the dispositions. Thus a magnet's disposition to attract iron filings is nothing more than the fact that its iron atoms are arranged in a structure that orients their electrons uniformly so that the laws of electromagnetism operate to create a magnetic field. Solubility of salt depends on chemical laws about sodium chloride bonds and the effects on them of polar molecules like those of water. Without laws about how the components of an object bring about its disposition, there are no dispositions!

These analyses make laws basic and dispositions into laws about the constituents of the things that have the dispositions. It's clear that Cartwright's theory reverses this order of conceptual dependence. It is the dispositions that things have which are basic. What we think of as a law is really a simplification that ignores all the other myriad dispositions a thing has, in order to explain its behavior to a good enough approximation. But really there are no laws; or maybe in physics there are a small number of fantastically complicated laws that accommodate every disposition of objects in its antecedent and so determine

the objects' actual path through space and time. But this "law" is so different from anything scientists could concern themselves with that we can ignore it for any scientific purpose or indeed any philosophical one. Moreover, such a complex law will inherit its nomic necessity from the necessary connection between having a disposition and manifesting it: being magnetic necessitates attracting iron filings.

Trading laws for dispositions may dispose of the problem of distinguishing laws from accidental regularities, but it has not relieved the philosopher of science from the responsibility of providing an account of necessity. Now it turns up as a relation between dispositions and the conditions under which they are manifested. This is something that several philosophers of science and metaphysicians are prepared to accept. These philosophers, like others freed from logical positivism's straightjacket prohibition of metaphysics, have widened the debate about nomic necessity to an extent that would be recognized by rationalists from the period before Kant wrote in the late eighteenth century.

Summary

It is hard to deny that scientists search for laws of nature, and that such laws are different from the accidental regularities that hold only temporarily and coincidentally in the history of the universe. Laws in fact are held to be different from, and "stronger" than, accidental regularities that might for all we know hold throughout the history of the universe. This fact about laws, that they seem to connect their antecedent to their consequents necessarily, needs to be explained. It's obvious that the task is one science does not address, but which needs to be discharged if we are to understand why laws explain.

The difference between laws and accidental regularities is reflected in the fact that laws are indispensable to deciding which counterfactual conditional statements we are prepared to accept as true and those we reject as false. But this role cannot explain the difference between laws and accidental regularities. Support for counterfactuals is a symptom of the difference that philosophy of science seeks to capture. Support for counterfactuals is also a mark of causal claims, whether in science or in everyday life. This suggests a connection between laws and causation that empiricists have claimed since Hume. Of course this connection only magnifies the seriousness of the problem of where physical or causal necessity may come from.

In this chapter we have canvassed several such explanations, as well as some attempts to explain away the difference, and even to explain away scientists' attachment to laws as mere heuristic devices. But the problem of how to understand physical or natural necessity cannot be avoided even by approaches that seem to deny that there are laws.

In fact, the entire debate about the nature of laws reveals just how much the problems of the philosophy of science recapitulate the fundamental problems that have absorbed philosophers since Plato: realism about laws goes back to Plato's dialogues, and its denial to Aristotle's famous work, *Metaphysics*, which

began the pursuit of the sub-discipline of philosophy named for it. The search for necessity in objects, among the local matters of particular fact, has been a pre-occupation of philosophers such as Locke and Leibniz, while others like Berkeley and Hume have argued that there could not be any such metaphysical glue in objects themselves. In later chapters we will again see how the understanding of science requires us to attend to these perennial issues in Western philosophy.

Study Questions

1. Most examples of laws of nature come from physics and chemistry. Explain why this is so.
2. According to the text, supporting counterfactual conditionals is a symptom of nomic necessity, not part of the explanation of it. Can an argument be made that counterfactual support is just what nomic necessity consists in?
3. Can we directly observe causation every time we see a pair of scissors cut or a hammer pound? If we can, what philosophical problems might this solve?
4. Which view of the nature of scientific explanation, the covering law approach or the pragmatic approach can dispense with the claim that laws have nomic necessity?
5. Few non-philosophers are Platonists. Many "scientific" philosophers are. Explain why.
6. Provide some arguments for realism about abstract objects.
7. If the laws of nature are true by definition, the problem of why they bear necessity is easy to deal with. Why is this and can such an argument be mounted?
8. Which of these concepts is harder to understand, "law of nature," "causation," or "counterfactual"? Does this have any implications for clarification of the meaning of concepts like law and causation?
8. Defend or criticize: "The problem of nomic necessity is a good example of the sort of pseudo-question that logical positivists sought to avoid. They were right to do so, as it can't be solved and doesn't need to be solved for any scientific purpose."

Suggested Readings

Marc Lange, *Natural Laws in Scientific Practice*, is a compendious and original examination of the nature of laws and their roles in all the sciences.

W. Kneale, *Probability and Induction*, advances a strong and influential account of the natural necessity of laws.

The role of counterfactual conditionals in understanding the nature and function of laws was first made clear in Nelson Goodman's *Fact, Fiction and Forecast*. An important contribution to the debate is J. L. Mackie's *Truth, Probability and Paradox*. One important chapter of this book, "The Logic of Conditionals," is anthologized by Balashov and Rosenberg. The most influential work on the logic of counterfactuals is David Lewis, *Counterfactuals*. Marc Lange, *Laws and Lawmakers*, argues for a frank metaphysics of subjunctive facts to explain laws' counterfactual support.

Hume advanced his theory of causation in Book I of *A Treatise of Human Nature*. Its influence in the philosophy of science cannot be overstated. *Hume and the Problem of Causation* by T. L. Beauchamp and Alex Rosenberg expounds and defends Hume's view. J. L. Mackie, *The Cement of the Universe*, provides an exceptionally lucid introduction to the issues surrounding causation, causal reasoning, laws, and counterfactuals, and defends an empiricist but non-Humean view.

A. J. Ayer, "What Is a Law of Nature?" reprinted in Curd and Cover, assimilates natural or nomic necessity to the way that we treat our best guesses about the laws of nature.

John Earman's best-systems account, "Laws of Nature," is anthologized by Balashov and Rosenberg. The latter-day origin of this view is in a paper written in the 1920s by Frank Ramsey, reprinted in a collection of his papers, *The Foundations of Mathematical Logic*.

Realist views of laws start with Fred Dretske, "Laws of Nature," and include D. M. Armstrong, "What Is a Law of Nature?" reprinted in Curd and Cover and in Lange's anthology. R. M. Tooley, *Causation: A Realist Approach*, presents a similar theory of causation.

William Lycan's *The Philosophy of Language* provides a cogent introduction of the theory of reference that seems to confer very strong, perhaps even logical, necessity on some laws of nature.

Nancy Cartwright's views begin with "Do the Laws of Physics State the Facts?" reprinted in Curd and Cover, and are further developed in her *How the Laws of Physics Lie*. This view is developed into a necessitarian treatment of dispositions as providing the nomic necessity in laws or even displacing them, by Chris Swoyer, "The Nature of Natural Laws," and more recently by Brian Ellis, *Scientific Essentialism*, and Alexander Bird, "The Dispositionalist Conception of Laws."

An excellent recent anthology on laws and their necessity is John Carroll's *Readings on Laws of Nature*.

5   Causation, Inexact Laws, and Statistical Probabilities

Overview

The covering law model makes laws indispensable to scientific explanation. Pragmatic accounts of scientific explanation that reject the covering law model recognize that in many contexts laws do explain. The previous chapter sought the explanatory power of laws in some kind of nomic or physical necessity, something they seem to share with causal sequences.

It is obvious, however, that many explanations, including ones deemed to be scientific, cite laws very different from those of physics or chemistry, or else don't cite laws at all. If we are to accept explanations in biology, the behavioral and social sciences, and history, as scientific—as providing knowledge—we at least are going to have to qualify the covering law model, or perhaps surrender it. There just don't seem to be enough laws for our explanatory needs, or at least not enough laws that are like those of physics. What is more we have seen that laws share with causes the same natural necessity that makes for scientific explanations. This makes a further exploration of causation and causal explanation urgent. Perhaps causation does the same work in other disciplines that strict laws do in so much physics and chemistry?

An examination of causal explanation makes it clear that what we identify as the cause of an event is almost always merely one among many

conditions that could bring it about, but by no means guarantees that it will happen. Often these causal explanations rest on or even include non-strict laws that connect causes and effects that don't always go together. These laws are non-strict because they contain *ceteris paribus*—other things being equal—clauses. Explanations that cite such laws or causes cannot satisfy the logical positivist requirement of giving enough grounds that we could have expected the explanandum event to have occurred. This may not be a problem if we forgo the demand that a good explanation could have served as a prediction.

But *ceteris paribus* laws are difficult to subject to empirical test: we can't ever be sure that "all other things are equal." Besides such qualified laws, there are ones that report probabilities, and these come in two varieties. Some statistical generalizations, like the one examined in Chapter 2, reflect limited knowledge and are stop-gap substitutes for strict laws. Others, like the basic laws of quantum physics, report ineliminably statistical disposition. But such non-epistemic probabilistic dispositions or capacities are difficult for empiricist philosophers of science to accept, for they do not appear to be grounded in more fundamental facts that could support these dispositions.

Some philosophers have therefore sought a feature of scientific explanation that is deeper than its employment of laws and commitment to reporting causal relations. They have grounded the nature of explanation in the unification of disparate phenomena under deductive systems that explanations, and especially explanations of laws, often provide.

Such an approach might respond to a traditional complaint made against scientific explanations. This is the notion that they tell us only how something happens, and not really why. Some who advance this view insist that the complete and final explanation of things will somehow reveal the intelligibility of the universe or show that the way things are is the only way that they could be. Historically famous attempts to show this necessity reflect a fundamentally different view of the nature of scientific knowledge than that which animates contemporary philosophy of science.

Causes as Explainers

The D-N or covering law model identified laws of nature as the source of science's explanatory power. Counterexamples to this model from within science suggest that we need to analyze scientific explanations as answers to questions, instead of deductive arguments. Our interest is in *scientific* explanations, as opposed to other kinds of (non-scientific) explanations. What we want to know is what makes them scientific: to use a term introduced in Chapter 3, we seek the "relevance relation" between a question and its explanatory answer(s), which will distinguish it from the pseudo-explanations of astrology or for that matter from the non-scientific explanations of history or everyday life.

We noted in Chapter 3 some of the suspicions that the logical positivists harbored about causality. Once those grounds began to seem more problematical than causation itself, philosophers of science turned their attention to causation. They did so for two reasons. First, as Hume noticed in the eighteenth century, causation seems to be "the cement of the universe"—the basic relationship between events, states, and processes that characterize reality. As we will see in this chapter, problems arose in twentieth-century physics, especially quantum mechanics, which seemed to require a clear understanding of causation. Second, many philosophers held that a direct focus on causality would be able to illuminate features of explanation, especially in the non-physical sciences, that the covering law model could not.

The claim that what makes an explanation scientific is the fact that it is causal goes back in some ways to Aristotle, who distinguished four different kinds of causes. Of these, the one that science has accepted as explanatory ever since Newton is the notion of an "efficient cause"—the immediately prior event that gives rise to, produces, or brings about what the explanandum describes. Physics seems to have no need for the other kinds of causes that Aristotle distinguished. This is because of physics' apparent commitment to mechanism—the thesis that all physical processes can be explained by the pushes and pulls exemplified when billiard balls collide. Biology and the human sciences apparently call upon a second type of causes that Aristotle identified, the so-called "final" causes—ends, goals, purposes—for the sake of which events occur. For example, it appears to be a truth of biology that green plants use chlorophyll *in order to* catalyze the production of starch. We will turn to final causes, and especially their role in biology and the human sciences, in the next chapter. For the moment, we will consider some of the problems that surround the notion of efficient cause that we need to deal with if causation is to illuminate scientific explanation.

The first of these problems was identified in Chapter 4: an account of the nature of causation must distinguish causal sequence from mere coincidence. The problem is the same as that facing the distinction between laws and accidental generalizations. It is all well and good to note that, like laws, causal statements support their counterfactuals, and do so because they express some kind of necessity. But we must not mistake the symptoms of causal necessity for its source. Indeed, it was the search for the source of causal necessity that led philosophers from Hume in the eighteenth century onward to look to laws of nature as the ground of causation. As we have seen, however, they face the same problem all over again. Perhaps we should reconsider the nature of causation, decoupling it from laws of nature? This strategy has been attractive for many reasons, among them the fact that scientists—especially social and behavioral scientists—provide causal explanations even though they have discovered few laws that compare with those of physics and chemistry for explanatory power and predictive reliability.

A second problem for efficient causes focuses on the actual character of causal explanations inside and outside science, which reveals their pragmatic

dimensions, their complicated relation to laws, and shows the difficulties of actually satisfying the D-N model or any analysis of scientific explanation like it. Suppose the lighting of a match is explained by citing its cause—that the match was struck. It is obvious that striking was not sufficient for the lighting. After all, had the match been wet, or there been a strong breeze, or no oxygen, or had the match been previously struck, or the chemical composition defective, or …, or …, or …, the match would not have lit. And there is no limit to the number of these qualifications. So, if striking was the cause, causes are at most only necessary conditions for their effects. But so are all the other qualifications concerning other necessary conditions—the presence of oxygen, the absence of dampness, the correct chemical composition, etc.

What then is it about a cause, as opposed to a mere condition, that enables it to explain? If causes and conditions are all just necessary conditions, why is the cause "the difference-maker," the trigger, the factor among all the others that answers the question, "why did the explanandum occur?" Some philosophers have argued that the context of inquiry is what makes this distinction: in the context of an evacuated chamber used to test match heads for hardness by striking them, the difference-maker of a match's lighting is not the striking, but the presence of oxygen (which should not be present in an evacuated chamber). Notice that this approach seems to make causal claims pragmatic or interest-relative. If we aim to uncover a relationship among events, states, and processes in the world that obtains independent of anyone's interests, inquiries, and other facts about them, instead of the world, this way of identifying difference-makers won't do. And if our further aim is to ground explanation on objective causal relations in the world, an account of causes that relativizes them to explanatory interests and background knowledge won't do either.

At this point we may be tempted simply to reject the metaphysical and even the scientific distinction between causes and merely necessary causal conditions. And we may begin to have a sneaking sympathy for the logical positivist's motivation in avoiding causality altogether. Or we may decide that causes are really sufficient conditions, and that the strict laws of physics will enable us to identify them in terms quite different from those that we use in the causal claims of ordinary life. None of these moves is available once we have agreed that sciences without laws can provide scientific explanations by citing causes.

In recent years a proposed solution to this problem has been widely discussed both by philosophers and some methodologists in the social sciences. James Woodward has argued for an account of causation that solves this and a number of other puzzles about scientific explanations. According to Woodward, C causes E if there is some "intervention" on C that does or would have changed E. One way that interventions happen is through human manipulation. These are not the only ways they can happen, but they are the ones we are most familiar with in everyday life. Strictly speaking, something is an intervention on C if the change in C is due to it, and the change in E comes about only owing to the changes in C. Notice that this approach is particularly natural when C and E can take on numerical values, i.e. are the sort of quantitative variables that interest

science and that are often subject to controlled experiments. In fact, Woodward's approach is inspired by the experimental scientist's interests and methods.

Examples of how interventions identify causes will help: pressing the brake pedal causes a car to slow down, because some intervention on the brake pedal or on the foot that presses that brake will result in the car's not slowing down, so braking slows the car down; preventing the needle on a barometer from going down will not result in dry weather, so the barometer's rising does not in fact cause the fine weather to persist. In these examples it is obvious that the intervention decides the matter. What Woodward's analysis does is show that across a wide variety of much more complicated cases, interventions identify causes (and disqualify putative causes). These are "objective" matters of fact about the interventions that enable us to identify causes and so underwrite or justify the explanations that cite them. This will be so even where we are ignorant of the relevant laws, or even in cases where there *are* no laws. By controlled experiments we can zero in on which manipulations or interventions on C are effective in changing E without even looking for, let alone discovering, any laws about Cs and Es.

Indeed, Woodward's approach deals with many objections and counter-examples to the covering law model, and shows why sometimes real scientific laws don't explain at all. One such classic objection to D-N explanation illustrates why. Consider the case of a male who takes birth-control pills daily. It is arguably a law that those who take birth-control pills do not become pregnant. The facts and the law together imply that the male does not become pregnant, but no one supposes that this explains it. If causes are identified by interventions, just get the male to stop taking the pills. Since pregnancy does not ensue, taking the pills couldn't have been the cause, and so doesn't explain it, even though it satisfies the covering law model's requirements.

The appeal to interventions in this approach to causation helps us to understand how laws provide explanations, but more important they help us understand why scientific explanations can be provided without appeal to laws. The most basic universal laws describe causal relations that are invariant under all interventions. No intervention will change the rate at which the gravitational force falls off with distance. It may sound funny but there is no way to prevent the speed of light from being the fastest speed at which anything can travel. But, if a regularity remains invariant under a broad enough range of interventions, then it may have explanatory power even if it is not a law. For example, the fact that robins' eggs are blue, together with the observation that an egg was laid by a robin, will explain why that particular egg is blue, even though there is no law that all robins' eggs are blue. One reason that this is not a law is that certain "interventions," such as the appearance of a new predator that differentially detected blue eggs, could falsify it: in such a case natural selection might eventually result in robins' eggs ceasing to be blue. But this particular intervention can reasonably be excluded from interventions like diet or climate that are relevant at the moment to whether being laid by a robin nowadays causally explains its egg's color.

An account of causation and of explanation without laws that exploits the notion of an intervention that brings about a change in the value of a variable raises some broad questions: First, it is clear that the notion of an intervention is one that cannot be understood without a prior commitment to the truth of counterfactual conditionals. This will be especially clear for interventions that human manipulation cannot bring about. To decide whether a change is an intervention where human intervention is not feasible, or feasible but not introduced, requires a decision on what would have been the case had the change not occurred. This of course means that the interventionist analysis of causation has already presupposed the sort of causal necessity of which counterfactual support is symptomatic.

The second question is even broader and a generalization of the first one. Woodward's intervention account of causation is itself steeped in causal concepts. For instance, an intervention on C brings about a change in E only if it does not directly cause a change in E—one that doesn't "go" directly through C. Thus, the theory is not what is often called a reductive analysis. A reductive analysis of causation is one that gives necessary and sufficient conditions for the truth of a statement of the form "C causes E," and each of the conditions provided can be understood without already understanding what causation itself is. Hume's analysis of causation is reductive: causation consists, according to Hume, in spatio-temporal contiguity (or contact), priority in time, and constant conjunction. A non-reductive analysis is open to the complaint that it is circular, or fails to resolve the problem which called for the analysis in the first place, or otherwise fails fully to elucidate the concept in question. In offering his interventionist account of causation, Woodward recognized that his analysis is not reductive, but insisted that it nevertheless provides an important understanding of how causal claims are grounded in science, how causal concepts are interrelated, and how scientific explanation proceeds.

Third, there is the problem of why it is that local regularities that are not invariant over every intervention nevertheless can provide explanations. One explanation for this fact that a proponent of the covering law model would give is that the local regularity is itself explained by an invariant general law operating on local conditions that gives rise to the locally invariant regularity: for example, the current distribution of the continents can be explained by regularities about plate tectonics in geology. These regularities, however, are just the result of much more fundamental laws of nature operating on the conditions of the Earth over several hundreds of millions of years. If this is correct, then explanations in terms of causal invariants and local regularities of limited invariance will turn out to be explanatory just because they implicitly appeal to completely invariant, strict laws of the sort we know of only in physics and perhaps chemistry. This result would be unfortunate for any theory of explanation that hopes to vindicate explanations in social and behavioral science even when they do not cite any laws. But this is certainly one of the chief aims of a theory like Woodward's.

This challenge makes it important to examine directly the question of whether the sort of local, limited regularities that figure in most of the life sciences along

with the social and behavioral sciences qualify as laws or not. This is not merely the honorific question of whether we should call them laws. It is the question of whether they carry explanatory force in spite of their exceptions, qualifications, statistical form, and recalcitrance to empirical testing. We return to this question in the next chapter. Meanwhile, our inquiry into whether causation can provide a basis for explanation has returned us to further questions about laws.

Ceteris Paribus Laws

If causes are difference-makers—necessary conditions of special sorts—then of course citing a cause will not by itself give good grounds to expect its effect. If we already know that the effect has taken place, as in most cases of explanation, this is not a problem. But when we engage in prediction of course we don't know this. Moreover, if we seek a complete explanation (assuming there is such a thing), or one that would enable us to reproduce the event to be explained in an experiment or in some technological application, merely knowing what the difference-maker on one occasion was may not suffice. We may also want to know the positive and negative conditions that together with the cause are needed to bring about the effect. We will need to know not just the necessary conditions, but the sufficient ones.

Now we can see one reason why the positivists preferred to appeal to laws rather than causes as the vehicles of explanation. A law of the form "All As are Bs" or "Whenever A occurs, B occurs," or "If A, then B" fulfills the good-grounds condition since its antecedent (A) is the sufficient condition for its consequent (B). However, if laws mention sufficient conditions for their consequents, and they underwrite causal sequences (as most philosophers of science hold), then these antecedents will have to include all the necessary conditions that, along with a cause, bring about its effect. For instance, a law about match-strikings being followed by match-lightings will have to include clauses mentioning all the conditions besides the striking that are individually necessary and jointly sufficient for match-lightings. If the number of such conditions is indefinitely large, the law cannot do this, at least not if it can be expressed in a sentence of finite length. This means either that there is no law of match-striking and -lighting, or that if there is, its antecedent includes some sort of blanket "other things being equal" or *ceteris paribus* clause to cover all of the unstated, indeed perhaps even unimagined, necessary conditions needed to make the antecedent sufficient for the lighting.

Of course there is no law about match-strikings and -lightings. Rather, the laws that connect the striking to the lighting are various, large in number, and mostly unknown to people who nevertheless causally explain lightings by appeal to strikings. This in fact will be the case for almost all explanations in ordinary affairs, and in the non-physical sciences, where laws have not (yet) been discovered, or where in fact there may be no laws of the sort familiar from physics. And explanations that don't cite such laws will not satisfy the positivist demand that it give good grounds to suppose that its explanandum actually occurred. It will be at most what positivists used to call an "explanation-sketch."

Even in those non-physical sciences that have claimed to uncover laws, explanations will not satisfy the positivist requirement built into the D-N model. For in these sciences, the laws will almost all contain implicit or explicit *ceteris paribus* clauses. They will take the form, "All other things being equal, if A then B," or "*Ceteris paribus*, whenever A then B." Such laws do not state the sufficient conditions for their consequents, and so can't guarantee that their consequents obtain, thus failing the adequacy condition for scientific explanations that the positivists laid down. Of course we may not accept this adequacy condition, but *ceteris paribus* laws have their own difficulties. As we shall see, such laws are hard to test, because their "other things equal" clauses excuse them whenever they appear to be disconfirmed. And these difficulties may even affect physical laws.

Recall, from the last chapter, Nancy Cartwright's argument that even Newton's laws contain implicit *ceteris paribus* clauses. For example, the inverse square law of gravitational attraction tells us that the force between two bodies varies inversely as the square of the distance between them. But we need to add a *ceteris paribus*—other things being equal—clause which rules out the presence of electrostatic or magnetic forces. Cartwright employed this argument to try to show that there are no strict laws in nature, only dispositions. Other philosophers have argued that even if Cartwright is correct about Newton's laws, this shows rather that they are not laws at all, just hypotheses that turned out to be false. The laws that have superseded them, for example, the laws of the general theory of relativity, or the standard model of the sub-atomic particles, really are strict laws that state sufficient conditions for their consequents.

Even if Cartwright is correct, there are only a small number of fundamental physical forces, so the problem of testing laws qualified by *ceteris paribus* clauses may be manageable in basic physics. But what happens when the number of conditions we need to hold constant increases greatly, as it does in biological generalizations (robins' eggs are blue), or those that figure in economics, for example (*ceteris paribus*, an increase in supply is followed by a decline in price)? Some philosophers have argued that each of the non-physical sciences is characterized by its own "proprietary" *ceteris paribus* laws that are explanatory in their intended domains. Following Jerry Fodor, philosophers often describe sciences with their own proprietary *ceteris paribus* laws as "special sciences." (More on such laws in the next chapter.)

The proprietary laws of the special sciences will be much the same as the explanatory regularities with restricted invariances that figure in Woodward's approach to causation. They will either be *ceteris paribus* laws or will implicitly appeal to them. But such non-strict laws raise problems for the philosophy of science, especially for its interest in demarcating science from non-scientific or pseudo-scientific explanations. As the number of possible interfering factors to be held constant grows, the testability of laws is reduced, which makes it too easy for anyone to claim to have uncovered a scientific law. This in turn threatens to trivialize causal explanation. If most of the laws we actually invoke in explanations carry implicit or explicit *ceteris paribus* clauses, then testing these laws requires establishing that other things indeed *are* equal.

But doing so for an inexhaustible list of conditions and qualifications is obviously impossible. And this means that it will be difficult to find differences in kind between real laws with inexhaustible *ceteris paribus* clauses, and pseudo-laws without real nomological (i.e. law-based) force—disguised definitions, astrological principles, New Age occult theories of pyramid power or crystal magic. These latter "laws" too can be protected from apparent disconfirmation by the presence of *ceteris paribus* clauses. "All Virgos are happy, *ceteris paribus*" cannot be disconfirmed by an unhappy person with a mid-August birthday since we cannot establish that besides the person's unhappiness, all other things are equal. This immunity from disconfirmation, along with wishful thinking, explains the persistence of astrology.

The testability of laws is something to which we will return at length in a later chapter, but there are consequences of this problem for our understanding of how science explains. In particular, when we exchange the appeal to causes for an appeal to laws, we avoid one problem—the relativity of causal judgments—at the cost of having to deal with another—the need to deal with *ceteris paribus* clauses. This problem is all the more pressing owing to a contemporary debate about whether there are any strict laws—exceptionless general truths without *ceteris paribus* laws—anywhere in science. If the inverse square law of gravitational attraction, for example, contains a proviso excusing counterexamples resulting from the operation of Coulomb's law in the case of highly charged but very small masses, then perhaps the only laws without *ceteris paribus* clauses to be found in science are the most fundamental laws to be found in the general theory of relativity, in quantum mechanics, or perhaps only in superstring theory.

Statistical Laws and Probabilistic Causes

In the biological and social sciences, instead of strict laws or even *ceteris paribus* laws, one finds statements of probabilities, or statistical regularities, and explanations that appeal to them. We encountered some of the problems that such explanations create for the D-N model in Chapter 3. Since many statistical generalizations seem to be explanatory, the question therefore arises of whether such statements are approximations to strict laws, under what conditions they may be understood to express causal relations that will explain particular phenomena, and most of all, what the nature is of the probabilities they express.

Explanations in medical contexts commonly employ epidemiological relations that are reported in statistical form, but are taken to express causal relations or at least to be the grounds for causal explanations. For example, it is widely accepted that smoking causes lung cancer because it is associated with a 12-fold increase in the probability of contracting lung cancer. For this reason we accept explanations of an individual's contracting lung cancer that the individual smoked, even though this may have only raised his or her statistical probability of contracting lung cancer from 0.013 to 0.17.

But we know full well that statistical correlation does not by itself warrant or reflect causal connection. What else do we require to move from statistical

correlation to causation? Along with this problem there is a further and equally serious one. We need to understand the nature of the probability at work in causal processes. To see the problem, consider another sort of statistical claim that is important in physics: to describe how events cause changes in probabilities. For instance, an electron passing through detector X will increase the probability that another one will pass through detector Y by 50 percent.

These two kinds of probabilistic causal claims are significantly different from one another. One is meant to be a statement in part about our knowledge; the other is a claim that is supposed to hold true even when we have learned everything there is to know about electrons. Each of them makes for a different problem in our understanding of causality.

There are two problems with saying that smoking causes cancer when the probability of a smoker contracting cancer is 0.17, and the probability of a nonsmoker doing so is 0.013. Some smokers never contract cancer, while some lung cancer victims never smoked. How do we reconcile these facts with the truth of the claim that smoking causes an increase in the probability of cancer? The fact that some lung cancer victims never smoked is not so serious a methodological problem. After all, two effects of the same kind can have quite different causes: a match may light as a result of being struck, or because another burning match was touched to it, or because it was heated to the kindling temperature of paper. The first fact, that some smokers don't contract lung cancer, is harder to reconcile with the claim that smoking causes cancer. Most smokers don't contract lung cancer at all. Even if we don't require constant conjunction for causation, we need to reconcile this fact with the causal claim that smoking does cause lung cancer. Obviously, there are other causally necessary conditions to be added.

So, what makes smoking the difference-maker that justifies our causal explanation and all those policies against smoking? One proposal that philosophers have made goes like this: Smoking can be said to cause cancer if and only if, among all the different background conditions we know about (heredity, diet, exercise, air pollution, etc.), there is no correlation between smoking and a lower than average incidence of lung cancer, and in one or more of these background conditions, smoking is correlated with a higher incidence in lung cancer rates.

Notice that this analysis relativizes causal claims to our knowledge of background conditions. Insofar as we seek a notion of causation that reflects relations among events, states, and processes independent of us and our theorizing about them, this analysis is unsatisfactory. But can't we just substitute "all background conditions" for "background conditions we know about"? That would eliminate the reference to us and our knowledge. Unfortunately, it also threatens to eliminate the probabilities that we are trying to understand. For "all background conditions" means the detailed, specific, causally relevant circumstances of each individual who smokes. And by the time we have refined these background conditions down to each individual, the chance of the individual contracting cancer will turn out to be either 0 or 1. If the underlying causal mechanism linking smoking and specific background conditions to cancer is deterministic, reflecting strict laws instead of probabilities, then our probabilistic causes will

disappear. The fact that causal statements based on probabilities reflect our available information will be a problem for the D-N model or any model that treats scientific explanation as a relation between statements independent of our beliefs. On the other hand, pragmatic accounts of scientific explanation will need to be filled in, as noted above, with conditions on what sort of information about statistical data makes an explanation scientific. We cannot accept an analysis of scientific explanation that makes scientifically relevant just anyone's answer to an explanatory question. The challenge is to identify the further epistemic conditions—requirements that need to be satisfied for statistical regularities to provide causal explanations, or any other kind of scientific explanation.

By contrast with probabilistic causal claims that seem to reflect limitations on our knowledge, there are what philosophers have called "objective chances"— probabilities whose values are raised and lowered by events in the world, quite independent of anyone's knowledge. The existence of these probabilities and the events that cause them is a consequence of the basic laws of physics. And these apparently exceptionless laws are ineradicably probabilistic.

Perhaps the most familiar is the second law of thermodynamics, which tells us that in any closed system, entropy will probably increase. Then there are laws like "the half-life of U^{235} is 6.5×10^9 years," which means that for any atom of U^{235} the probability that it will have decayed into an atom of lead after 6.5×10^9 years is 0.5. Laws like these do not merely substitute for our ignorance, nor will they be replaced through refinement to strict non-probabilistic ones. Quantum mechanics tells us that the fundamental laws operating at the basic level of phenomena are just brute statements of probability, which no further scientific discoveries will enable anyone to reduce or eliminate in favor of deterministic strict laws. The law about the half-life of uranium attributes to uranium atoms an "objective chance" of decaying. Being composed of a certain number of electrons, protons, and neutrons causes a uranium atom to have a certain probability of decaying in the next minute. This probability remains constant over time and is the same for every uranium atom, and there is no difference at all between two uranium atoms, one of which decays in the next minute and another one that doesn't. The objective chance or probability of decay is a quantitative measurable tendency, a disposition, a **probabilistic propensity** to decay at a certain rate. But the objective chances in these laws present us with still another difficulty for causation. The causal probabilities of quantum mechanics are "tendencies," "dispositions," "capacities," "propensities" or powers of some sub-atomic arrangements to give rise to certain effects.

These probabilistic powers are troublesome to some scientists and many philosophers. This is because, as we noted in Chapter 4, many philosophers think that dispositions can really only be understood by explaining them in terms of further, more basic non-dispositions. To see why they think this, consider a non-probabilistic disposition, say, fragility.

A glass is fragile if and only if were it to be struck with sufficient force it would shatter. But, note, this is a counterfactual statement, and it will be accepted only if there is a law that supports it—a law that reports a causal relationship between

glass being fragile and shattering when struck. And this law about fragile objects obtains owing to a causal relationship between the molecular structure of glass and its shattering when struck. All (normal) glasses are fragile but many glasses never shatter. Their fragility consists in their having the molecular structure reported in the law that supports the counterfactual. In general, attributing a disposition, or capacity, or power to something is tantamount to hypothesizing a causal relationship between some of that thing's non-dispositional, structural properties and its behavior. Being fragile is having a certain structure, a structure that the object has all the time, even when it is not being struck or shattering. Here is another example: A piece of metal's being magnetic is a matter of attracting iron filings, and its being a magnet consists in the arrangement of atoms that make it up in a lattice, and the orientation of the electrons in these atoms. But this arrangement is present in a magnet even when it is not exerting a magnetic force on anything nearby.

Applying this result to the probabilistic propensities of thermodynamics and quantum mechanics reports is problematic. Physicists and philosophers have been trying to ground the probability reported in the second law, about the probable increase of entropy, on more fundamental non-dispositional and non-probabilistic facts about matter in motion ever since the law was first stated in the nineteenth century. This attempt is in accordance with the empiricist conviction that dispositions must be grounded in occurrent, manifest, actual structures. So far they have not succeeded.

The problem of grounding propensities in quantum physics is even more serious. Since these probabilities are propensities or dispositions, and are the most fundamental, "basement"-level properties physics reports, there cannot be a more basic level of structural properties to causally ground these probabilities. They are therefore "free-floating" powers of microphysical systems, which the systems probabilistically manifest, but which when not manifested still exist without any actual causal basis. Compare fragility or magnetism: Can these potentials be present in a glass or a piece of iron without some actual property to underlie them—such as molecular composition, or orientation of outer-shell electrons in a lattice? No. Without such a "base" it is difficult to understand probabilistic propensities as dispositions, powers, or capacities with causal foundations. We cannot establish their existence as distinct from their effects, the frequencies with which the quantum effects they bring about occur. There is nothing to show for them independent of our need to somehow ground probabilistic regularities at the basement level of physics. These pure probabilistic dispositions will be very different from the rest of the dispositional causes that science cites to explain effects. Unlike fragility or magnetism or any other disposition that science studies, quantum probabilistic propensities are beyond the reach of empirical detection (direct or indirect) independent of the particular effects that they have. They have all of the metaphysical mystery of the concept of causal or nomological necessity and more besides—for the objective chances of quantum mechanics support counterfactuals and provide explanations, and could do neither without being physically necessary.

Explanation as Unification

We have now explored some of the problems that must be addressed by those who seek to ground scientific explanation on the notion of causation. It may now be easier to see why many philosophers, and not just logical positivists, have hoped to find an analysis of the nature of explanation in science that avoids having to face such intractable questions about the nature of causality, or laws—strict or not.

It may be worthwhile to step back from the project of analyzing causal explanation, and explanation by the invocation of laws, to ask a more basic question: What does scientific understanding consist in? If we have an answer to that question, it may help us identify what it is among causal explanations and nomological explanations that enables them to convey scientific understanding. One answer to the question of what sort of understanding we seek in providing scientific explanations goes back at least to an insight of Albert Einstein's, according to which scientific theorizing should "aim at complete coordination with the totality of sense-experience" and "the greatest possible scarcity of their logically independent elements (basic concepts and axioms)." The demand for "scarcity" is translated into a search for unification.

On this approach a scientific explanation conveys understanding because it affects unifications, which reduces the stock of beliefs we need to have in order to effect explanations. The two key ideas are these: First, that scientific explanations should reflect the derivation of the more specific from the more general, so that the stock of basic beliefs we need is as small as possible. Second, which stock of basic beliefs we embrace is constrained by the need to systematize experiences. Unification is the aim of scientific explanation, because, on this view, human understanding of the world increases as the number of independent **explanantia** we need decreases. So, in the explanation of general phenomena, what makes an explanation scientific is that phenomena are shown to be special cases of one or more general processes; in the explanation of particular events, states, and conditions, what makes for scientific explanation is that the explanans on the one hand apply widely to other explananda, and that the explanans themselves be unified with other beliefs by being shown to be special cases of other more general explanans. According to Philip Kitcher—who is one of the chief proponents of this view of scientific explanation—the demand for unification makes logical deduction an especially important feature of scientific explanations. Indeed, this is what unification consists in. (We shall return to the role of deduction in explanation when we examine the nature of theories in Chapter 7.) Kitcher also requires that the propositions that affect unification pass stringent empirical tests. These two conditions on unification show that this alternative still shares important similarities with the D-N model of explanation. And they reflect the sort of thinking that stands behind the best-systems view about the nature of nomological necessity. These shared characteristics are among its strengths, yet it purports to go deeper than Hempel's general criterion of adequacy (that the explanans give good grounds to expect the explanandum)

and systems maximizing simplicity and strength to some underlying feature of scientific explanation.

Unification does seem to contribute to understanding. But let us ask, why? What makes a more compendious set of beliefs about nature better than a less compendious set, assuming that both account for the evidence—data, observations, experiences, etc.—equally well? One answer might be that the universe is simple—that the underlying causal processes that give rise to all phenomena are small in number. In that case, the search for unifications will reduce to the search for causes, and strict laws that connect them to their effects. If causation is, as empiricists have long held, a matter of laws of increasing generality, and if the universe reflects a hierarchy of more basic and more derived causal sequences, then explanations that affect unification will also uncover the causal and nomic structure of the world. When it has done so it will need to address our questions about how causes and laws explain. It will not have avoided them.

Now, suppose that the universe's causal structure is permanently hidden from us, because it is too complex or too submicroscopic, or because causal forces operate too fast for us to measure or are too strong for us to discern. Suppose further that we nevertheless can effect belief-unifications that enable us to systematize our experiences, to predict and control up to levels of accuracy good enough for all our practical purposes. In that case, for all its practical payoff, unification will not enhance understanding of the way the world works, or will do so only up to certain limits.

Proponents of unification may have a more philosophically tendentious argument for distinguishing unification from causation, and preferring it. They may hold, along with other philosophers of science, that beyond observations the causal structure of the world is unknowable and so drops out as an epistemically relevant criterion on the adequacy of explanations. More radically, they may hold (as Kitcher does) that causation consists in explanation, or that causation, like explanation, is also unification-dependent. So, unification is all that scientific understanding can aim at. We will return to these issues later (in Chapter 7), in our discussion of the nature of theories.

Summary

If causes explain then they have to be different from other conditions, which like causes must be present to bring about the effect that they explain. How to draw this distinction between causes and mere conditions independent of an analysis of explanation is a problem that must be solved if we hope to use causation to illuminate explanation.

The way that *ceteris paribus* laws seem to work in explanations suggests that this can be done, though without showing how. What is more, these laws, so widespread in the social and behavioral sciences, are difficult to test and too easy to defend. This robs them to some extent of their explanatory power.

Statistical regularities often replace these *ceteris paribus* laws and seem to be an improvement on them, at least in testability. Yet in the way they rely on and

reflect our ignorance of all the causal factors that bear on an outcome to be explained, they too seem only weakly explanatory. By contrast, the probabilistic causation of fundamental physics has all the strength that exceptionless strict laws can provide yet raises several of the gravest metaphysical questions that already face physical law, but now in new ways.

These problems of causal and nomological explanation have suggested to some philosophers that there must be a more fundamental feature that all explanations share in common, which enables them to enhance understanding. One version of this idea has been developed as the claim that scientific explanation works by unification, which reduces the number of basic beliefs we must have about the world to enable us to unify as much of its apparently disparate phenomena as possible. This is an idea that will resonate in several ways in the chapters to come.

Study Questions

1. If, as some philosophers argue, all laws have *ceteris paribus* clauses, what implications are there for limits to explanation, and to prediction?
2. What sorts of regularities have greater explanatory power, inexact *ceteris paribus* laws or precise statistical regularities?
3. Why is it difficult for empiricists to accept quantum mechanical probabilities as fundamental unexplainable facts about the world?
4. How different is the D-N model from the view that scientific explanation is a matter of unifying disparate phenomena?
5. Why is deciding on the meaning and significance of *ceteris paribus* clauses crucial to the methods of the social sciences?
6. Should science aim to reduce or eliminate the role of statistical probabilities in its explanations and theories?

Suggested Readings

Aristotle advanced his theory of four causes in the *Physics*. The problem of *ceteris paribus* clauses is treated insightfully in one of Hempel's last papers, "Provisos," reprinted in A. Grunbaum and W. Salmon, *The Limitations of Deductivism*. Nancy Cartwright, *How the Laws of Physics Lie*, is the *locus classicus* for arguments that all laws bear *ceteris paribus* clauses. An excellent discussion is Marc Lange, "Who's Afraid of Ceteris-Paribus Laws?", reprinted in his anthology. John Earman and John Robert argue that "There Is No Problem of Provisos."

Important recent works on causation and causal explanation are James Woodward, *Making Things Happen*, C. Glymour, P. Spirtes, and R. Scheines, *Causation, Prediction, and Search*, and Michael Strevens, *Depth*. These books

are heavily influenced by the work of Judea Pearl, including *Causality: Models, Reasoning, and Inference*, and *The Book of Why: The New Science of Cause and Effect* (written with Dana Mackenzie).

J. L. Mackie, *Truth, Probability, and Paradox: Studies in Philosophical Logic*, includes two exceptionally clear essays from an empiricist perspective on the meaning of probability statements and on the problem of dispositions.

Kitcher expounds his account of explanation as unification in "Explanatory Unification and the Causal Structure of the World," anthologized in Balashov and Rosenberg. The original exposition of this view is to be found in W. Salmon and P. Kitcher, *Scientific Explanation*, as well as a paper anthologized in Pitt, *Theories of Explanation*. This anthology also contains a paper developing the same view independently by M. Friedman. Wesley Salmon's critique of the unification account and defense of a causal view of explanation is developed in "Scientific Explanation, Causation, and Unification," reprinted in Balashov and Rosenberg

Salmon has long been particularly concerned with statistical explanation, a matter treated along with other topics in his *Scientific Explanation and the Causal Structure of the World*. Salmon's own views are expounded in "Scientific Explanation, Causation, and Unification," reprinted in Balashov and Rosenberg, as is Kitcher's defense of explanation as unification, "Explanatory Unification and the Causal Structure of the World."

6   Laws and Explanations in Biology and the "Special Sciences"

Overview

Beyond causation and unification, people have sought even more from scientific explanations: purpose and intelligibility. Both the explanation of human action and of biological processes proceed by citing their purposes or goals (people work in order to earn money, the heart beats in order to circulate the blood). On the one hand, these explanations don't seem to be causal; after all, the explanans (the ends or goals) obtain after the explanandum (the means to attain the goal) in these cases. On the other hand, purposive explanations in biology and the human sciences seem more satisfying than explanations in physics. Moreover, physical science has ruled out future causation, even if common sense ever allowed for its possibility. So, whether and how these "**teleological**"—goal-directed—explanations can be reconciled with anything like causal explanation remains to be addressed.

The appeal of teleological and interpretative explanations, in contrast to causal ones, brings us face to face with a challenge alleged to face science in general: it only explains the *how* of things, and never really *why* they happen. This traditional complaint that scientific explanation does not really tell us why things happen comes with an expectation that the

complete and final explanation of things will somehow reveal the intelligibility of the universe or show that the way things are is the only way they could be. Historically famous attempts to show this necessity reflect a fundamentally different view of the nature of scientific knowledge from that which animates contemporary philosophy of science.

Dissatisfaction with Causal Explanations

Whether scientific explanation is causal, unificatory, nomological, statistical, deductive, inductive, or any combination of these, a question remains about how and whether scientific explanations really convey the sort of understanding that satisfies our inquiry. One longstanding perspective suggests that scientific explanation is limited, and in the end unsatisfying, because it does not get to the bottom of things. Sometimes this perspective expresses itself in the thesis that scientific explanations reveal only *how* things come about, but not *why* they happen. Thus, for example, it will be held that all a D-N model tells us about an explanandum-event is that it happened because such an event always happens under certain conditions and these conditions obtained. But when we want to know *why* something has happened, we typically already know *that* it has, and may even know that events like it always happen under certain circumstances. What we want is some deeper insight into why it came about.

When this sort of dissatisfaction with scientific explanation is expressed, what sort of explanation is sought? These deeper explanatory demands seek an account of things that shows them (and nature in general) to be "intelligible"—to make sense and add up to something, instead of just revealing a pattern of one damned thing after another. Traditionally, there seem to be two sorts of explanations that aim at satisfying this need for deeper understanding than the push–pull, "efficient" cause explanations that physics and chemistry provide.

Sometimes, the demand is for an explanation which will show that what happened *had* to happen in a very strong sense—that its occurrence was necessary, not just physically but as a matter of rational intelligibility or logic. Such an explanation would reveal why things couldn't have turned out any other way than the way that they did. It might, for example, reveal that the laws of nature are not contingently true about the world, but necessarily true—that there is only one possible way for the world to be. On this view, gravity cannot, as a matter of logical necessity, vary as the cube of the distance between objects as opposed to the square, copper must as a matter of logic alone be a solid at room temperature, the speed of light couldn't be 100 miles an hour greater than it is, etc. As noted in Chapter 1, this is a conception of science that goes back to the eighteenth-century rationalist philosophers Leibniz and Kant, who set themselves the task of showing that the most fundamental scientific theories of their day were not just true, but necessarily true, and thus provided the most complete form of understanding possible.

There is a second sort of explanatory strategy that seeks to respond to the sense that causal explanations are unsatisfying. This goes back much further than the eighteenth-century philosophers, back at least to Aristotle, who clearly identified the kind of explanatory strategy in question. This is the notion of "final-cause" explanations that appear to be common in biology, the social and behavioral sciences, history, and everyday life. In these contexts, descriptions and explanations proceed by identifying the end, goal, or purpose, for the sake of which something happens. Thus, green plants have chlorophyll *in order to* produce starch, Caesar crossed the Rubicon *in order to* signal his contempt for the Roman Senate, the central bank raised interest rates *in order to* curb inflation. In each of these cases, the explanation proceeds by identifying an effect "aimed at" by the explanandum-event, state or process that explains it. These explanations are called "teleological," from the Greek *telos*, meaning end, goal, purpose. There is something extremely natural and satisfying about this form of explanation. Because it seems to satisfy our untutored explanatory interests, it may be thought to serve as a model for all explanation. To the extent that non-purposive explanations fail to provide the same degree of explanatory satisfaction, they are sometimes stigmatized as incomplete or otherwise inadequate. They don't give us the kind of "why" that final-cause, purposive explanations do.

The attractions of an explanation that shows why the thing that happened *had* to happen as a matter of logical necessity, and the appeal of teleological explanations, are based on very controversial philosophical theses, which most philosophers have repudiated. If these two sorts of explanation rest on questionable assumptions, then it will turn out that despite the feeling that it isn't enough, "efficient" causal explanation will be the best science or any other intellectual endeavor can offer to assuage explanatory curiosity.

Teleological explanations seem to explain causes in terms of their effects. For example, the heart's beating—which is caused by its circulating the blood—explains the effect. Since the time of Newton such explanations have been suspect to philosophers and physicists. In the words of the seventeenth-century philosopher Spinoza, they "reverse the order of nature," making the later event—the effect—explain the earlier event—the cause. The problems that teleology poses are multiple and obvious. If future events do not yet exist, then they cannot be responsible for bringing about earlier events. Physics does not allow for causal forces (or anything else for that matter) to travel backwards in time. This prohibition is entrenched in the special theory of relativity that forbids causal processes traveling faster than light's speed. And any process moving from the future to the past would have to be much faster than that! Moreover, sometimes a goal that explains its cause is never attained: producing starch explains the presence of chlorophyll even when the absence of CO_2 prevents the green plant from using chlorophyll to produce starch. Most metaphysical theories disallow non-existent events, states and processes from bringing about actually existing ones. Thus, physical theory and perhaps metaphysics conspire to make things difficult for future purposes to bring about the past events, states and processes that are the means to attain them, and so to explain them.

The exclusion of teleology from physical processes raises three possibilities: First, if physics does not allow "final causes," then there are none and those to be found in biology and the human sciences are just illusory pseudo-explanations. Or, second, biological and other apparently teleological processes are irreducibly different from physical processes. Or, third, despite their appearance, when we understand how they really work, teleological processes are not so different from efficient causal processes. On this third alternative, once we understand how teleological processes work, we will discover that they are just complicated causal processes. This will turn apparently teleological explanations into implicit causal and nomological explanations no different in kind from those that physics accepts and employs.

The first two alternatives are philosophically controversial: it seems hard to deny that some things in nature (at least us) have purposes. Allowing biology to have purposes when physics has ruled them out is mysterious to say the least. So, the third alternative is worth exploring. Can explanations that appear to appeal to purposes really turn out to be garden variety causal explanations of the same type as physics employs?

Proprietary Laws in the "Special Sciences"

Explanations of human action in everyday life, in history, biography and human affairs, appear to be teleological. They proceed by identifying people's goals, ends, plans, and purposes, and then explain the actions by showing that they were means to those goals and purposes. Because these explanations appear to cite future ends to account for events—human actions in particular—that occur earlier and are means to these ends, some philosophers have argued that explanations in everyday life, history, and the human sciences are not causal at all. They are only meant to make human events "intelligible." They are "interpretative," and explain events by redescribing them in light of the outcomes they aim to achieve. If this is a correct account of the nature of explanation in the human sciences, their explanatory strategies will be utterly different from those of natural science. The contrast between explanations that cite causes and those that confer intelligibility is part of a widely employed argument that aims to exempt the social sciences and history from honoring methodological rules common to the natural sciences.

Despite such appearances many philosophers of science hold that these explanations are not really teleological at all, and that through conferring some special intelligibility to human affairs, they are still causal explanations. It's just that these causal explanations identify desires and beliefs as the causes that bring about those actions that they also make intelligible. These explanations only look teleological because the desires and beliefs are about future states or events, and are identified in terms of them. Thus my buying a ticket for Friday's London to Paris train on Monday is explained by the desire I had on Monday to go to Paris on Friday. There is no future causation here, though there is a description of the prior cause—the desire felt on Monday to bring about its future effect, my going to Paris on Friday.

If these explanations are causal, then many philosophers will hold that there must be a law or regularities of some kind that link desires and beliefs on the one hand as causes to their actions as effects. It is not difficult to locate them in the explanations of human action given in the social and behavioral sciences.

Many explanations and theories in social science presuppose that there are laws of rational choice. These take the form of variations on the following three statements:

1. Between any two choices facing an individual, e.g. going to Paris or going to Brussels, the person prefers one to the other or is indifferent between them. (comparability)
2. If *a* is preferred by the individual to *b*, and *b* to *c*, then the agent prefers *a* to *c*. (transitivity)
3. The individual always chooses the most preferred among available alternatives. (preference maximization)

Sometimes these three statements are treated as parts of the definition of a rational agent or of rationality:

> An individual is rational, by definition, if the individual satisfies comparability, transitivity, and preference maximization.

If these three conditions define what it is to be rational, they cannot serve as laws that explain actual choices. But, as we will see in Chapter 9, the definition of rationality can and often is employed as a model that can be exemplified, to a greater or lesser extent, by people under various circumstances. Economists seek to apply rational-choice models wherever exchange occurs. In effect, they adopt as a factual, empirical, contingent hypothesis the generalization that actual choosers satisfy the model of rationality closely enough for it to be useful in understanding their actual choices

Of course when treated as factual generalizations (and not parts of a definition) none of the three statements about choice is true, nor is the generalization about all choosers. The only way they can be defended as laws is by assuming that they have *ceteris paribus*—other things being equal—clauses attached to them. It is a widespread view that the so-called "proprietary laws" of the "special sciences" take this form.

The expression "**special sciences**" usually refers to the social and behavioral sciences. The label "special" suggests that these disciplines neither uncover nor employ universal laws like those of physics but regularities that obtain only in special circumstances, such as when life emerges in the case of biology, or when exchange emerges as in the case of economics. Laws about the kinds of things each special science focuses on are "proprietary" in the sense that the special science "owns" and "operates" the law; it has the default authority to assert its explanatory power in the discipline's domain, and its instances constitute the domain of the sciences that own it.

Treating the three conditions for rationality as a model is equivalent to adopting such a *ceteris paribus* law. Nowadays, many of the special sciences are devoted to framing such models. This is why they face the problems raised for such non-strict laws even when they do not explicitly frame them. We noted in Chapter 5 the difficulties of testing *ceteris paribus* laws. Their role in demarcating the domain of the sciences in which they are to be found makes this problem even more serious. Behavior that does not conform to the proprietary law of a special science may all too easily be excluded or ignored in a special science, instead of taken to disconfirm the law and undermine explanations that employ it.

If we deny that there are such *ceteris paribus* laws in the special sciences, we must either deny that their explanations are scientific or find another source for their explanatory power. The first alternative is implausible. Few wish to deny that the special sciences provide any explanations of events in their domains of inquiry. A few philosophers have held that the laws that underwrite explanations in these disciplines are to be found in neuroscience or in biology. The fact that social scientists explicitly deny the relevance of these laws to explanations in the special sciences undermines this approach. This leaves the view that explanation in the special sciences is not causal, nor otherwise based on laws, but is an enterprise different in kind from natural science. Perhaps desires and beliefs bring about or cause actions, but that is not how they explain them. Rather they confer understanding by providing the meaning of the actions, interpreting them, and somehow making them intelligible to us. This is why explanations in the human sciences, the humanities, history, and common sense are satisfying in a way that causal explanations never could be.

As noted, this view is widely held by social scientists and others who reject the relevance of methodological rules that are extracted from physics and imposed on them. These philosophers reject causal analysis of beliefs, desires, hopes, fears, and other psychological factors that explain human action. Causation cannot confer intelligibility and these explanations do just that. Yet this approach raises the question of what intelligibility is.

There are thus at least two fundamental issues to be faced in the philosophy of the "special sciences." First, what kinds of things are the mental states that explain action? This is a problem that goes back at least to Descartes' argument in the seventeenth century for "dualism," the thesis that the mind is not the brain and mental states are not physical ones. Second, what is the source of the intelligibility mental states allegedly provide the human events they explain? The first question (the mind–body problem) is a fundamental issue in metaphysics and we cannot treat it in any depth here. The second problem of whether explanations need to or can confer intelligibility, and if so how, is pursued further at the end of this chapter.

Functional Laws and Biological Explanations

Explanations in the social sciences seem to be purposive, goal-directed, teleological. This is a feature that such explanations share with biological explanations.

Compare: Why did Caesar cross the Rubicon? In order to overthrow the Roman republic; Why does the heart pump the blood? In order to circulate oxygen to the body. If intelligibility has its sources in teleology, then an examination of how explanations proceed in biology may help illuminate the source of this alleged intelligibility. After all, teleological explanations in biology will be simpler than in the human sciences; they will not involve psychological processes. So we will be able to separate the matter of how teleological explanations work from the problems of how mental and physical states are related.

Empiricist philosophers of science will welcome this approach in part because they are interested in the nature of biological explanation. But they also believe that an analysis of it will vindicate their claim that all scientific explanation is ultimately causal or law-governed; that there is no special intelligibility that explanation in the special sciences confers that is lacking in physical science.

It's ironic but obvious that until the middle of the nineteenth century teleological explanations in biology were treated as belief/desire explanations of the sort still common in the human sciences. In this case the beliefs and desires that did the explaining were those of God—the omnipotent and benevolent deity who designed everything in the biological realm. What is more, until the nineteenth century the hypothesis that crucial facts about organisms were to be explained in this particularly satisfying way—by appealing to God—seemed a reasonable one.

Before Darwin's theory of natural selection, arguably the likeliest explanation for the adaptedness and complexity of biological organization was provided by citing God's design: one rendered biological organization intelligible by giving the purpose that the organism's parts play, their roles in God's plan for the survival and flourishing of the organism. Thus, why does the heart pump the blood? The answer "in order to circulate oxygen" is an abbreviation for something like "God in his goodness wanted to arrange for blood circulation, and unerringly believed that the best way to do this was to build hearts. Since he was omniscient, he was correct in this belief, and since he was omnipotent, he was able to act on it. That is why hearts pump blood."

Scientific explanation is no longer permitted to appeal to a deity, so an entirely different analysis of teleological explanations is required. What is more, as noted in Chapter 1, with the advent of Darwin's theory of evolution the need for such theological explanations in biology was completely eliminated. Darwin showed that adaptation—the appearance of purpose or design—always results from a purely causal process of heritable variation that is blind to adaptive needs, and natural selection that filters out the less well-adapted heritable variations. A fuller exposition of how Darwin's theory does this is given in Chapter 9. What that presentation will make clear is that the appearance of design is the result of purely causal processes in which no one's purposes, goals, ends, intentions, etc., will play any actual part. Thus, green plants bear chlorophyll because at some point (through blind variation) their predecessors happened to synthesize some chlorophyll molecules, the endowment was inherited, and since chlorophyll happens to catalyze the production of starch, producing starch kept these

plants alive longer and they had more offspring. Randomly produced increases in the amount of chlorophyll synthesized resulted in still more offspring, which out-competed plants lacking chlorophyll until the only plants left were plants with higher concentrations of this molecule. And that's why latter-day plants have chlorophyll.

The "in-order-to" of our original explanation gets cashed in for an etiology in which the filter of natural selection culls those plants that lacked chlorophyll or its chemical precursors, and selects for those which have it, or mutated from the precursor closer and closer to chlorophyll as it exists in present-day green plants. And where do the first precursor molecules come from, on which nature selects and selects and selects until chlorophyll emerges? That first precursor is the result of purely non-foresighted chemical processes, to be explained by chemistry without recourse to its adaptive significance for plants.

This strategy has been advocated by philosophers as the best way to analyze all teleological explanations in natural, social, and behavioral science, not just botany. All apparently teleological explanations answer the question "Why X?" or "Why does X happen?" The answer takes the form, "X occurs in order that Y occurs" and these answers are always to be analyzed into an "etiology," a sequence of past events in which anything that brought about Y-type events in the past gets selected for. Usually the selection is natural and the etiology involves genetic heredity, which ensures the persistence and adaptational improvement of the precursor Xs selected for just because they bring about Y or its precursors. So, "Hearts pump the blood in order to circulate oxygen" is short for something like this: hearts in fact circulate oxygen because in the evolutionary past of creatures with hearts there was selection for those genes that build oxygen circulators, and this persistent selection filtered for mutations and combinations of genes that fortuitously built better muscle proteins and the enzymes that catalyzed their composition into cardiac muscles and packages of them, until at last the process reached the level of adaptation to be found in current vertebrates. There is not supposed to be anything teleological left in this "translation" into purely cause/effect language of what apparently teleological explanations and regularities say. All the purpose has been cashed in for causation. For obvious reasons this purely causal account of purpose in terms of a consequence etiology is called the "selected-effects" analysis.

In biology this analysis of teleological laws and explanations is attractive, but it faces some difficulties. For example it makes the truth of claims about the current function or adaptation of some biological trait turn on facts that occurred literally geological ages ago. And if some "preadaptation" comes to be co-opted to a new use, through a relatively recent environmental change, the selected-effects view will have difficulty. In this case the trait won't be present owing to selection on its current function, but arguably on its prior function. This may in fact be true for most complex adaptations. Some of the objections to this analysis of purpose can be accommodated in various ways. But the selected-effects approach has deeper problems, especially outside biology, in other disciplines that employ teleological language.

The special sciences are rife with teleological explanations, descriptions, and regularities. In social science, **functional** analyses are widespread. For example, consider a statement such as "The function of the price-system is to communicate information that people need in order to make productive and efficient decisions about what to buy and to sell." This claim cannot be denied and is clearly teleological: the price system exists *in order* to meet a human need. But it was not designed by anyone to do so, so we cannot identify a prior etiology that gave rise to it through variation and selection of beliefs and desires of people who wanted to meet their needs. Still less is the price system in our genes! Neither natural selection nor human intention in the past is available to cash in many of the functional claims of the social sciences for purely causal ones. The same problem arises in the life sciences. Anatomists and physiologists frequently advance hypotheses about what parts and components do that enable the larger system to perform in its characteristic way. These hypotheses about what parts do are claims about their functions. But the researchers offering these claims reject any commitment to an evolutionary etiology as part of the meaning of what they claim, or even as required for their claims to be true. In 1921 Banting and Best first identified the function of the isles of Langerhans in the pancreas— to produce insulin that digests glucose in the blood—by experiments on dogs. This led to the first effective treatment for diabetes. They were not, however, engaged in any sort of evolutionary inquiry nor had they any interest in the "etiology" of the biological process of digestion. By discovering the function of these cells, they were isolating their *causal role* in digestion.

These problems led to a quite different approach to functional and teleological explanations, descriptions, and regularities in the special sciences (named by its originator, Robert Cummins, as the "causal-role" account of function). The first thing that this approach insists on is that when scientists claim that "The function of X is to do Y" their interest is rarely if ever to explain the presence of X by giving its purpose. If the blood-pumping function of the heart is not actually cited by scientists anymore (if it ever was) to explain why vertebrates have hearts, there is no need to search for a causal etiology of past events to take away the apparent suggestion of future causation this statement seems to make. Discovering functions and attributing them are the stock in trade of many of the special sciences. But if scientists do not explain the presence of a trait by identifying its function (and so buying into a consequence etiology for it), what are they doing? Something quite different: they are analyzing a complex behavior away into its simpler components.

The causal-role theory claim that X has a function F is an abbreviation for a more complicated statement roughly of this form: X's having F as an effect contributes to a complex capacity of some large system that contains X as a part, and there is an "analytical account" of how X's F-ing contributes to the "programmed manifestation" of the complex capacity of the containing system. Thus X's F-ing is a "nested capacity"—a less complicated capacity that along with different capacities of other parts of a system that contains X, enables the entire system to discharge some relatively more complicated behavior. Notice

how this analysis ties into an approach to laws examined at the end of the previous chapter. It treats capacities and dispositions as more basic than any strict laws. Of course, functional or teleological laws such as that hearts beat in order to pump blood will have to contain implicit *ceteris paribus* clauses if they are to be proprietary laws of one or another special science. The problem of such laws might be held to disappear or to be solved if the proprietary laws of each of these disciplines were actually expressions of the basic dispositions of objects in the intended domain of the special sciences.

Various strategies are available to reconcile the selected-effects analysis of purposive and functional descriptions with one another. Even if they cannot be reconciled, they both still have serious consequences for any suggestion that the special sciences do not trade in causal explanations, or that their explanations carry some special non-causal explanatory force. As we have seen, both accounts of functional or apparently purposive explanations in biology, and the other special sciences, trade in teleology for ordinary causal processes. As such they break down significant objections to the relevance of methods from physical science in biology, and the relevance of methods from the natural sciences, including biology, to the social and behavioral, or "special" sciences.

Explaining Purposes or Explaining Them Away?

Darwin's revolutionary accomplishments have a relevance to every aspect of intellectual debate in contemporary culture. Their implications for religion are so obvious that even 150 years after the publication of *On the Origin of Species*, Darwin's theory is controverted by those who rightly fear its inimical implications for theism.

For 2,500 years after astronomy got started among the Babylonians, the realm of living things was deemed to be unexplainable by the resources of physics. The great eighteenth-century philosopher Immanuel Kant argued that "there will never be a Newton for the blade of grass." Within 60 years of this pronouncement, the Newton of biology (Darwin) had succeeded in making purposive explanation safe for natural science, showing how purposes are just complex combinations of causal sequences, along with their unusual causal histories.

Darwin's achievement has sometimes been given an alternative interpretation. Instead of holding that he made purposes scientifically acceptable, one could argue that he rid nature of purpose. After all, what Darwin did was to show how a purely causal process—blind variation and environmental filtering ("natural selection")—can produce adaptations, biological structures with functions in a "causal role." In doing this he revealed that the appearance of design and purpose was mere illusion—an overlay that we rolled out on a purely mechanistic world. The appearance of purpose had led scientists and almost everyone else to accept the existence of a deity, designing and executing a plan as the only explanation for appearances. But now we not only see that no such designing deity is required, we also see that the appearance of purpose is just that, appearance and not reality.

On this interpretation of Darwin's achievement, we should conclude not just that the appearance of design was produced without its reality, but that there is no deity whose plan gives rise to the adaptation and complexity of biological systems. We may go on to infer there is no meaning—nor any real intelligibility—to be found in the universe, or at least none put into it by anyone but us. There may remain room in the scientist's ontology for a deist's conception of God as the first cause, but no room for cosmic meanings endowed by God's interventions in the course of nature.

But whether Darwin expunged purpose from nature or naturalized it, one thing he certainly did was show that in explaining biological phenomena we need not appeal either to God's prior intentions or to forces from the future giving rise to adaptations in the past or present.

The hope therefore must be surrendered that teleology or purpose—prior design or causes operating from the future to bring about events that lead up to them—could be the source of a kind of intelligibility transcending the understanding provided by causal explanation. Those who seek a source for understanding, scientific or otherwise, beyond causation and contingent laws will not find it in biology's final causes.

From Intelligibility to Necessity

We are left with the first of our two sources of dissatisfaction with causal explanation: the idea that it does not provide for the intelligibility of nature.

How could it do this? Those who reject the explanatory methods of empirical science because they fail to go beyond causes, laws, or unifications will not find intelligibility in final causes or human interpretation. These are not really alternatives to what natural science provides. Explanations that appeal to purpose and function turn out in the end to be false or disguised variants of causal explanations. Interpretations may satisfy the itch of curiosity but they are untestable and unreliable in prediction and control. We cannot mistake the fact that interpretation does satisfy human curiosity for the claim that it provides deeper understanding than science, or even any understanding at all. For any non-empirical, non-experimental discipline to do that would require the vindication of some intelligible version of rationalism regarding empirical topics that philosophy has struggled to offer.

Since the seventeenth century there have been rationalists with such an agenda. They have held the view that scientific explanation should uncover underlying mechanisms responsible for the course of nature that reveal that there is only one course that nature could have taken. Two important eighteenth-century philosophers—Leibniz and Kant—argued that science does in fact reveal such necessities. As such, science's explanations (when complete) leave nothing unexplained, allow for no alternative account, and therefore bear the highest degree of adequacy.

Leibniz sought to show that through physical knowledge we would see that each law fitted together with the rest of scientific theory so tightly that a change

in one law would unravel the whole structure of scientific theory. The inverse square law of gravitational attraction could not have been an inverse cube law without some other law having been different, and differences in that law would make for further differences in other laws until we discovered that the entire package of laws governing nature needed to be changed in order to preserve it from logical contradiction and incoherence. Hence, the package of all the laws will make one another mandatory or, as it were, internally logically necessary. This would confer a kind of logical inevitability to the way in which the course of the laws of nature has played out. Leibniz did not argue for this view by showing exactly how changes in our best scientific theories would ramify throughout the web of science. He could not do so because scientific knowledge was in his time too incomplete even to try. Yet it is still too incomplete to show any such incoherence under changes in its parts. Moreover, even if we acquired a package of scientific laws that work together to explain all phenomena, we would still need some assurance that this is the only package of scientific laws that would do so. The logical consistency of all of our scientific laws—indeed their arrangement in a deductive order that unifies them in a logical system—is by itself insufficient to rule out the existence of another such system, with different axioms and theorems, which might effect the same systematization of phenomena. This is the problem of "underdetermination," to which we shall turn in Chapter 12.

Interestingly, Leibniz solved the problem of multiple packages of internally coherent laws by appealing to teleology. He argued that among all the packages of complete systems of laws, logically so related that none could be revised unless every other law was changed, God chose the "best" of them to govern the actual world owing to his benevolence. For that reason the laws that govern phenomena in the actual world not only logically support one another, but the entire package is the only possible set of laws. So, if we accept Leibniz's confidence in divine benevolence, we will see that nomological explanations confer a very strong necessity on their explanantia (plural of explanans). Of course, if we are unprepared to help ourselves to divine teleology to underwrite every correct scientific explanation, we cannot share Leibniz's confidence in deductive-nomological explanations as reflecting either necessity or intelligibility.

By contrast to Leibniz, Kant was unwilling to appeal to God's intentions to underwrite science. But like Leibniz he was strongly committed not only to the view that scientific explanation had to reveal the necessity of its explanantia, but also to the claim that the scientific laws that Newton had discovered were the necessary truths to which physics at any rate had to appeal. Kant attempted to craft arguments to reveal the necessary truths at the foundations of Newtonian mechanics. His theory holds that the nature of space and time, the existence of a cause for every physical event—causal determinism—and, for example, the Newtonian principle of the conservation of matter, are necessary because they reflect the only way in which cognitive agents like us can organize our experiences. As such these principles can be known *a priori*—independently of experience, observation, experiment—through the mind's reflection on its own powers—its "pure reason." Whence the title of Kant's great work, *The Critique*

of Pure Reason. Unlike Leibniz, Kant recognized that scientific laws are not logical truths. By contrast with the laws of logic, and with statements true by definition, like "All bachelors are unmarried males," the denial of a scientific law is not self-contradictory.

Recall the discussion of Kant's approach to the nature of science discussed in Chapter 2. Kant held that scientific laws, whose denials are not self-contradictory, are synthetic truths, by contrast with analytic truths. Kant defined analytic truths as ones whose subject "contains the predicate." This is obviously a metaphor, but the idea is that analytic truths are true by definition or the consequences thereof. As Kant held, long before the logical positivists, analytical truths (as definitions or their deductive consequences) are without content, make no claims about the world, and merely indicate our stipulations and conventions about how we will use certain noises and inscriptions. For example "density equals the quotient of mass and volume" makes no claim about the world. It does not imply that there is anything that has mass, volume, or density. The definition cannot explain any fact about the world, except perhaps facts about how we use certain noises and inscriptions. If "having a certain density" could explain why something has a certain mass-to-volume ratio, it would be a case of "self-explanation"—an event, state, or condition explaining its own occurrence. For having a certain density just is having a certain mass-to-volume ratio. If nothing can explain itself, analytical truths have no explanatory power. A synthetic truth, by contrast, has content, makes claims about more than one distinct thing or property in the world, and thus can actually explain why things are the way they are. The laws of nature are thus synthetic truths.

Kant accepted that Newton's laws were universal truths and that they were necessary as well. Since he held that universality and necessity are the marks of *a priori* truths, Kant set out to explain how it is possible for the fundamental laws of nature to be "synthetic *a priori* truths." That is, they can make explanatory claims about the actual world even though we can know this fact about them and the world without recourse to observation, experiment, the collection of data, or other sensory experiences of the world. If Kant's program of establishing the synthetic *a priori* character of, say, physics, had succeeded, then its explanations would have a special force beyond simply telling us that what happens here and now does so because, elsewhere and elsewhen, events of the same kind happen in circumstances of the kind that obtain here and now. According to Kant, the special force that such law-governed explanations bear consists in these being the only laws that our minds can by their very nature understand, and their truth is assured to us by the nature of human thought itself. Pretty clearly, explanations of this character will be particularly satisfying, not to say exhaustive and exclusive of alternatives.

Kant believed that unless he could establish the synthetic *a priori* truth of at least physics, it would be open to skeptical challenge by those who deny that humans can discover natural laws. In particular Kant was concerned to refute an argument he identified as David Hume's: If the laws of nature are not knowable *a priori*, then they can only be known on the basis of our experience. Experience,

however, can provide only a finite amount of evidence for a law. Since laws make claims that (if true) are true everywhere and always, it follows that their claims outrun any amount of evidence that we could provide for them. Consequently, scientific laws are at best uncertain hypotheses, and the claims of physics will be forever open to skeptical doubt. Moreover, Kant feared (rightly, it turned out) that speculative metaphysics would inevitably seek to fill this skeptical vacuum.

Kant was correct in holding that the laws of nature are synthetic. However, for the philosophy of science, the most significant problem facing Kant's account of Newtonian theory as synthetic truths known *a priori* is that the theory isn't true at all, and so cannot be known *a priori*. What is more, its falsity was established as the result of experiment and observation. And since these experiments and observations underwrite theories—notably Einstein's theories of relativity, and quantum mechanics—that are incompatible with Newton's theory, then neither Newton's laws nor their successors could in fact be known *a priori*. Philosophers of science concluded that the only statements that we can know *a priori* will be those that lack empirical content, i.e. definitions and the logical consequences of definitions that do not constrain the world at all, and so have no explanatory relevance to what actually happens. Since experience, observation, experiment, etc. can never establish the necessity of any proposition, scientific claims that have any explanatory relevance to the way the world actually is cannot therefore be necessary truths. Two important consequences follow from this conclusion. First, it is a mistake to search for an alternative to causal explanation that reveals the necessity or intelligibility of the way things are, for necessary truths have no explanatory power. Second, if a proposition has explanatory power, then it must be a statement with content, in Kant's term synthetic, and not analytic. But such statements can only be justified by observation, experiment, and the collection of data.

This conclusion, however, leaves us confronted with the threat of skepticism that Kant's predecessor Hume recognized: since the empirical evidence for any general law will always be incomplete, we can never be certain of the truth of any of our scientific laws. Hume's argument is widely taken to show at least that science is inevitably fallible. If Hume is right the conclusions of scientific investigation can never acquire the sort of necessity required by Kant, Leibniz, and others who have craved certainty or necessity. But this fallibility will be unavoidable in any body of scientific laws that have explanatory content, which make claims about the way the world works. No one should expect logical necessity from science. Neither should they want it.

Summary

Scientific explanation has traditionally been met with dissatisfaction by those who demand that such explanation show the purpose, design, or meaning of natural processes, and not just how they came to happen. This demand for final-cause or teleological explanation goes back to Aristotle. Contemporary accounts of teleological explanation in biology exploit Darwin's discovery of

how blind variation and natural selection can give rise to the appearance of purpose. Darwin's theory helps us see that teleological explanation is only a complex and disguised form of causal explanation. Where functional explanations in the life sciences are not so understood, this is because they play an entirely different role, elucidating complex processes by identifying the causal contributions—the functions—that parts of a larger system play in delivering complex behavior.

Whether we can deal with such apparently teleological explanations in the human sciences in a similar way turns on whether and how people's cognitive and emotional attitudes explain their behavior, causally, or in some other way that exempts the human sciences from the methods of the natural ones. Social scientists and some philosophers have long sought such an exemption because the interpretation of human action seems to convey an intelligibility that causal explanation in natural science lacks.

Is there any basis for expecting or demanding something more from scientific explanation than the identification of contingent causes, that would make its explanations more compelling? There is a tradition, which goes back at least to the eighteenth-century philosophers Leibniz and Kant, of arguing that scientific explanation must ultimately show that science's description of reality is not just true, but necessarily true—that the way the world is, is the only way it could be. We have good reason to think that any attempt to establish such a conclusion is bound to fail. Indeed, were it to succeed, we would be hard-pressed to explain much of the fallible and self-correcting character of scientific knowledge.

Study Questions

1. Defend or criticize: "The fact that scientific explanation cannot provide for the intelligibility or necessity of things is a good reason to seek it elsewhere."
2. Does the Darwinian theory of natural selection show that there is no such thing as purpose in nature or does it show that there are purposes and they are perfectly natural, causal processes?
3. Is there any way to combine the Darwinian natural-selection approach to biological function with the causal-role approach due to Cummins?
4. Are there important differences between models and inexact *ceteris paribus* laws in the special sciences?
5. In what sense is biology, and all of the social and behavioral sciences, historical?

Suggested Readings

The way in which Darwinian theory can be used to assimilate purpose and teleology to causation is most influentially explained in L. Wright, *Teleological Explanation*. An anthology, C. Allen, M. Bekoff, and G. Lauder, *Nature's Purposes*, brings together almost all of the important papers on this central topic in the philosophy of biology. Cummins' work is anthologized and developed by others in Ariew, Cummins, and Perlman (eds.), *Functions: New Essays in the Philosophy of Psychology and Biology*.

Robert Brandon and Daniel McShea, *Biology's First Law*, argues for a fundamental law in biology that is clearly not teleological.

The nature of intentional explanation in the social sciences is treated in A. Rosenberg, *Philosophy of Social Science*.

Wesley Salmon, "Probabilistic Causality" is reprinted in Lange's anthology, as is Nancy Cartwright's "Causal Laws and Effective Strategies."

For more on models in the special sciences, see Chapter 9.

Jerry Fodor first argued for *ceteris paribus* laws in "Special Sciences: Or the Disunity of Science as a Working Hypothesis," reprinted in Curd and Cover. He revisits his argument in "Special Sciences: Still Autonomous After All These Years."

Much of Leibniz's work remains untranslated and what is available is very difficult. Perhaps most valuable to read in the present connection is his *New Essays on Human Understanding*. Immanuel Kant, *The Critique of Pure Reason*, defends the claim that the most fundamental scientific theories are synthetic truths known *a priori*. Hume's account of empirical knowledge is to be found in his *Inquiry Concerning Human Understanding*, which also develops Hume's account of causation.

7 The Structure of Scientific Theories

Overview

How often have you heard someone's opinion written off with the statement, "That's just a theory"? Somehow in ordinary English the term "theory" has come to mean a piece of rank speculation or at most a hypothesis still open to serious doubt, or for which there is as yet insufficient evidence. This usage is oddly at variance with the meaning of the term as scientists use it. Among scientists, far from suggesting tentativeness or uncertainty, the term is often used to describe an established subdiscipline in which there are widely accepted laws, methods, applications, and foundations. Thus, economists talk of "game theory" and physicists of "quantum theory," biologists use the term "evolutionary theory" almost synonymously with evolutionary biology, and "learning theory" among psychologists comports many different hypotheses about a variety of well-established phenomena. Besides its use to name a whole area of inquiry, in science "theory" also means a body of explanatory hypotheses for which there is strong empirical support.

But how exactly a theory provides such explanatory systematization of disparate phenomena is a question we need to answer. Philosophers of science have long held that theories explain because, like Euclid's geometry or Newtonian mechanics, they are deductively organized systems. It should be no surprise to learn that a proponent of the D-N model of explanation should be attracted by this view. After all, on the D-N model, explanation

is deduction, and theories are more fundamental explanations of general processes. But unlike deductive systems in mathematics, scientific theories are sets of hypotheses, which are tested by logically deriving observable consequences from them. If these consequences are observed, in experiment or other data collection, then the hypotheses are tentatively accepted. This view of the relationship between scientific theorizing and scientific testing is known as "**hypothetico-deductivism**." It is closely associated with the treatment of theories as deductive systems, as we shall see.

Our treatment of the nature of theories and how they work will proceed through the study of a particularly important theory, Newtonian mechanics. We employ this theory to illuminate questions about theory in general. But we also show why in many respects it effected a complete sea change in Western civilization's conception of the universe and our place within it.

How Do Theories Work? The Example of Newtonian Mechanics

What is distinctive about a theory is that it goes beyond the explanation of particular phenomena to account for more general phenomena. When particular phenomena are explained by an empirical generalization, a theory will go on to explain why the generalization obtains, and any exceptions to the generalizations under which it fails to obtain. When a number of generalizations are uncovered about the phenomena in a domain of enquiry, a theory may emerge that enables us to understand the diversity of generalizations as reflecting the operation of a single or small number of processes. Theories, in short, unify, and they do so almost always by going beyond, beneath, and behind the phenomena that empirical regularities report, to identify underlying processes that account for the phenomena we observe. This is probably the source of the notion that what makes an explanation scientific is the unifications it brings about. For theories are our most powerful explainers, and they operate by bringing diverse phenomena under a small number of fundamental assumptions.

For the philosophy of science the first question about theories is, How do they effect their unifications? How exactly do the parts of a theory work together to explain a diversity of different phenomena? One answer has been traditional in science and philosophy since the time of Euclid. Indeed, it is modeled on Euclid's own presentation of geometry. Like almost all mathematicians and scientists before the twentieth century, Euclid held geometry to be the science of space and his *Elements* to constitute a theory about the relations among points, lines, and surfaces in space.

Euclid's theory is an axiomatic system. That is, it consists of a small set of postulates or axioms—propositions not proved in the axiom system but assumed to be true within the system—and a large number of theorems derived from the axioms by deduction in accordance with the rules of logic. Besides the axioms

and theorems there are definitions of terms, such as a straight line—nowadays usually defined as the shortest distance between two points—and the circle—the locus of points equidistant from a given point. The definitions of course are composed from terms not defined in the axiomatic system, like point and distance. If every term in the theory were defined, the number of definitions would be endless, so some terms will have to be undefined or "primitive."

It is critical to bear in mind that an axiom that is assumed to be true in one axiomatic system may well be a theorem derived from other assumptions in another axiomatic system, or it may be justified independently of any other system whatever. Indeed, one set of logically related statements can be organized in more than one axiomatic system, and the same statement might be an axiom in one system and a theorem in another. Which axiomatic system one chooses in a case like this cannot therefore be decided by considerations of logic. In the case of Euclid's five axioms, the choice reflects the desire to adopt the simplest statements that would enable us conveniently to derive certain particularly important further statements as theorems. Euclid's axioms have always been accepted as so evidently true that it was safe to develop geometry from them. But, strictly speaking, to call a statement an axiom is not to commit oneself to its truth, but simply to identify its role in a deductive system.

It is clear how Euclid's five axioms work together to systematize an indefinitely large number of different general truths as logically derived theorems. For example, given the axioms of Euclid, you can prove that the alternative interior angles of a line between two parallel lines are equal. Then, using that theorem together with the definition of a straight line as a 180° angle, you can prove that a triangle has internal angles that add up to 180°. The Pythagorean theorem that the square of the hypotenuse of a right triangle equals the sum of the squares of the other two sides can also be provided from the axioms using other theorems. Many secondary school students used to spend an entire year deriving theorems from Euclid's axioms. Some students may still.

Perhaps the most famous theory in physics to be presented as an axiomatic system was Newton's mechanics. The theory originally consisted of three axioms, to which Newton later added a fourth of great importance. We will have occasion to mention them repeatedly hereafter, so it is a good idea to introduce them now, in a way that shows how they are related and can therefore work together.

Newton's first law was initially articulated explicitly by Galileo, and indeed, Newton gave him credit for discovering it. The law is deceptively simple, and obvious when you think about motion and force in the right way. Though it stared every physicist in the face for 2,000 years, it remained undiscovered until Galileo and Descartes came close to hitting upon it. More important, it effected the most profound break in the history of science. Thereafter it became a model for revolutionary change in science. The law simply states that:

1. An object moving at constant speed in a straight line continues to do so (forever) unless a force is applied to it.

The first thing to notice is that this law contradicts the common-sense view that to keep a body in motion requires the continual or repeated application of force. This notion is not only common sense, confirmed continually in our experience, but was the cornerstone of physics from Aristotle to the seventeenth century. Yet all you need to see that common sense is wrong and that Newton's first law is right is a simple thought experiment of Galileo's: Set a ball in motion down an incline plane. What does it do? It speeds up. Set it in motion up an incline plane and it slows down. Now, set it in motion on a perfectly flat, frictionless surface, what will it do? It can't slow down; it can't speed up. There is no alternative. It must continue to move with constant speed. We need to correct common sense. In particular we need to recognize that, in our experience, friction is never zero and that's why we have to apply forces to keep things in motion.

Why is this law revolutionary? Because it radically changes our notion of what it means to be "at rest." Science often takes the condition of being at rest as not requiring explanation: when nothing is happening, there is nothing to explain. Before Newton, physics treated "rest" as the condition a thing was in when its velocity was zero. So, when things moved they were not at rest, and their motion had to be explained. The explanation had to be that some force was acting on them. There is a mistake in this line of thinking, and it prevented physics from correctly explaining all motion, including the motion of the planets, for 2,000 years. Newton changed the definition of rest. In his theory, rest is not defined as the state of zero velocity, but as the state of zero acceleration. This slight conceptual change means that things in motion could be at rest. Therefore, their motion need not be explained, certainly not by seeking forces responsible for the motion. This is crucial: If things traveling at non-zero velocity but zero acceleration are not subject to any force, but your theory says they are, you will never find the forces your theory insists are present. If your definition of rest is zero velocity, you will never even be able to disconfirm the physics that is steering you wrong. Ask yourself, is common sense Newtonian or Aristotelian? Most people admit it's Aristotelian, and not just because it violates Newton's first law. The difference is a conceptual one, a matter of definition hard even to notice. Seeing things in motion as really being at rest is difficult, but we must do so, for anything else is an impediment to the advance of science.

Newton's second law may seem obvious once you understand the conceptual breakthrough of the first law. If bodies traveling at constant speed in a straight line are at rest, and no forces are acting on them, then when their speed is changing and/or they are not traveling on straight lines, but on any other path, then there must be some force acting on them. Right? This seems a natural thought, and it is the second law of motion:

2. The force acting on a body is equal to the product of its mass and its acceleration.

    $F = ma$.

According to this law, if something is moving on a curved path with constant speed, there must be some force acting on it. If something is moving on a straight line at changing speed, there must be some force acting on it. Neither of these cases constitutes rest, so both require explanation by uncovering the force acting in each case. Newton's real departure from previous physics, in this law, was the concept of mass. Mass is neither bulk nor weight, and Newton could say no more about it than that it was a "quantity of matter" and proportional to force. Newton's departure is hidden by the deceptive simplicity of the equation.

The third law in Newton's theory is the one that most obviously evokes the corpuscularian metaphysics of seventeenth-century philosophers and scientists. These philosophers rejected the idea that physical processes can have non-physical causes or that physical changes can bring about their effects "at a distance" across empty space. Newton bought into this view and articulated it in this law:

3.   For every action, every application of force, there is always an opposed equal and opposite reaction, a force applied in the opposite direction.

Thus when one object hits another and imposes a force on it, the other resists, imposing a force on the first object; for the same reason, to leave the ground an object at rest must exert a thrusting force against the Earth. Of all Newton's laws, this one seems the least departure from common sense. But its acceptance makes Newton's fourth law, the inverse square law of gravitational attraction, as problematic as it is powerful.

The key ideas in Newton's first three laws work together to explain many, many regularities. But perhaps their most impressive achievement is the way they led Newton to his most distinctive contribution to physics—the concept of gravity and the (independent) law that expresses its character.

Recall that the second law demands that forces are present when bodies move on curved paths. The planets obviously move around the Sun on curved paths and the Moon moves around the Earth on such a path. Therefore, they must be under the influence of some force or other. Similarly, though undetectable by us, the Earth is moving around the Moon, and so must be acted upon by forces as well. Even if we can't detect this force, the third law tells us that if the Earth is exerting a force on the Moon then the Moon must be exerting a force on the Earth. These forces cannot be contact forces of the sort that make the third law so appealing to corpuscularian philosophy, but they are forces nonetheless.

Another thing we know from experience is that more massive objects are heavier, and we need to exert more force, to do more work, to hold them up than less massive ones. Astronomy reveals that planets travel around the Sun in ellipses. The ellipses closer to the Sun are smaller, tighter, and more curved than the ones farther away. So, invoking the second law again, the force exerted between the Sun and closer planets must be stronger than that exerted between it and more distant planets. All these considerations work together to suggest a fourth law, the inverse square law of gravitational attraction, which Newton introduced well after the first three laws of motion:

4. The force between two objects is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. In symbols:

$$F = G\, m_1 m_2 / d^2,$$

where G is the gravitational constant equal to about 6.67×10^{-11} cubic meters per kilogram per second squared.

   As we will see, the force governed by this equation, the gravitational force, is quite different from the "contact forces" that dictate the third law. But we know that there has to be such a "non-contact" force, given the first and especially the second law. In fact, Newton's fourth law commits us to just the sort of force that the corpuscularian philosophy sought to expunge from physics! As a result, on the one hand, Newtonian mechanics became the most explanatorily compelling and predictively powerful theory ever discovered (and remained so until its successors, the general theory of relativity and quantum mechanics). But, on the other hand, its explanatory and predictive powers came with a commitment that Newton himself and every physicist who followed him was profoundly uncomfortable with: a commitment to the existence of a very "spooky" force that travels at infinite speed through complete vacuums and from which no shielding is possible. All three of these features, its speed, its ability to move through totally empty space, and its ability to penetrate any barrier, together made gravity different from everything else physicists thought they understood, on the basis of the corpuscularianism of their time. This mystery continued to haunt physics right into the twentieth century, long after corpuscularianism was strongly confirmed as the right view about everything except gravity. We will return to this problem in the next chapter.

   The laws of Newtonian physics "work together" not only because they are about the same things—mass, velocity, acceleration, and force—but because thinking about each of them helps us, and perhaps helped Newton, formulate the others. Besides the lines of thought mentioned above, there is another interesting relationship among Newton's laws worth mentioning, because of the light it sheds on laws in general. Given the inverse square law of universal gravitation, it turns out that every body in the universe is under the influence of gravitational forces from everything else. Consider what this means for Newton's first law: it has no positive instances or examples. There is not a single object in the universe moving in the absence of forces acting upon it. Yet, it is still a law: it still expresses some kind of nomological necessity even though it is "vacuously" true—i.e. has no instances because given the fourth law's truth, it can't have any. It is worth considering what implications (if any) this fact about Newton's first law has for the nature of laws and physical necessity in general. It is also worth considering whether the first law might not be better phrased as a law about how things behave when the "net" forces they are subject to cancel one another out. So interpreted it may have many instances and not be vacuously true after all.

More important, what makes these four laws one theory is that they work together to explain a vast array of other phenomena, some already well known in Newton's day, some discovered since then, and a good many discovered only by the use of Newton's laws as instruments of discovery.

The first thing Newton was able to explain as a direct consequence of his laws were Kepler's laws of the motion of the planets: that they travel in ellipses, and that the planets sweep out arcs of equal area in equal units of time. He was also able to explain most of Galileo's discoveries in mechanics, including his discoveries about free-falling objects having (almost) constant acceleration, the trajectories of cannon balls, and the behavior of pendulums, incline planes, levers, fulcrums, pulleys, etc. All this Newton was able to do by mathematical demonstration, by deduction from his four axioms. Subsequent developments in physics enabled his successors to do the same for regularities about the motions of comets, stars and ultimately galaxies, work and energy and its conservation, the behavior of solids, liquids, buoyancy, and hydraulics generally, aerodynamics, and as we shall see, eventually a great deal of thermodynamics. By the end of the nineteenth century, about the only physical processes that seemed beyond the reach of Newtonian explanation were the interrelated phenomena of light, magnetism, and electricity. And all of this explanatory power was revealed by mathematical deduction of the vast range of regularities from the underived four axioms of Newtonian theory.

Theories as Explainers: The Hypothetico-Deductive Model

The fact that Newton's four laws are axioms from which a number of other regularities in nature can be derived is not enough by itself to make them into one theory. What makes one theory out of several separate laws cannot merely consist in their logical relations to one another or even in the fact that they jointly imply regularities that no one of them could imply alone. There has to be something more to being a theory than merely having an axiomatic structure from which theorems can be derived.

To see the problem, consider the following "theory" composed of two axioms "working together" and the theorems deduced from them:

> The ideal gas law: $PV = nRT$, where P = pressure, T = temperature and V = volume, and r is the universal gas constant.
> The quantity theory of money: $MV = PT$, where M is the quantity of money in an economy, V = the velocity of money—the number of times it changes hands, P is the average price of goods, and T is the total volume of trade.

From the conjunction of these two laws, either one of them follows logically, by the simple principle that if "A and B" then "A." But so do other generalizations. For example, from $PV = nRT$ and some other definitions it follows that, when the pressure on the outside of a balloon is constant, an increase in temperature

increases its volume. From the quantity theory of money it follows that, other things being equal, increasing the amount of money in circulation results in inflation. Now we can put both of these consequences together into a fatuous regularity that heating a balloon by putting a candle under it and heating an economy by increasing its money supply increases the balloon's volume and the economy's price level. Notice both the ideal gas law and the quantity theory of money are required to derive this last regularity. Yet no one thinks there is a theory composed of these two "laws."

In a theory, the parts must work together to explain. But working together cannot be captured by the notion of logical derivation alone. Yet saying exactly what it is about the components of a theory that make it one theory, instead of several joined together, is the beginning of another longstanding philosophical challenge. For the philosopher of science it is not enough to say that a theory is a body of laws that work together to explain. "Works together" is too vague and, as we shall see, "logically implies laws as theorems" is too precise. More importantly, philosophers of science seek to clarify what it is about a theory that enables it to do the scientific work that it does—explaining a large number of empirical regularities, and their exceptions, and enabling us to predict outcomes to greater degrees of precision than the individual laws that the theory subsumes.

One natural suggestion might emerge from the conclusion of the previous chapters. The fundamental, underived general laws of a theory work together by revealing the *causal structure* of underlying processes. Laws governing these structures rise to be laws that the theory systematizes and explains. So, what's wrong with the theory composed of the ideal gas laws and the quantity theory of money is that there is no single underlying structure common to both the behavior of gases and money for there to be a theory about. How do we know this? Presumably because we know enough about gases and money to know that they have nothing directly to do with one another.

But even concepts like underlying causal structure or mechanism may not provide the degree of illumination we seek. Our previous discussion uncovered some serious reasons why philosophers are reluctant to place too much weight on the notion of causation. What is worse, the notion of an underlying mechanism may seem disturbing given the empiricist argument that there is nothing to causation beyond regular sequence; there is no glue, no mechanism, no secret power or necessity in nature to link events together in ways that make the course of things inevitable or intelligible. With these reminders about difficulties ahead and behind, we must nevertheless explore the notion that a theory is a body of laws that work together to explain phenomena by attributing an underlying causal structure or mechanism to the phenomena. We must do so because so many theories manifestly work like this.

Newton's theory and its development provide an illuminating example of how the basic laws of a theory work together to explain empirical regularities. By the late eighteenth century it was apparent that Newton's laws did not seem to govern the behavior of solid objects. This departure from what the theory predicted could not be written off as the result of friction. But the great mathematician and

physicist Euler was able to show that in fact Newton's theory *could* explain and predict the behavior of three-dimensional solids that are subject to deformations and other "real-world" effects. To do so, one assumes that such bodies are composed of Newtonian corpuscles that themselves cannot be deformed but that do behave exactly in accordance with Newton's laws. The laws governing the actual behavior of extended bodies that are derivable from Newton's laws operating on their constituents unsurprisingly are known as Euler's laws. This is the sort of development of Newton's theory that secured more and more confidence as the eighteenth and nineteenth centuries wore on.

Another, and perhaps the philosopher's favorite, example of how Newton's axioms work together as a theory is provided by the kinetic theory of gases. The development of this theory illustrates very nicely many different aspects of theoretical advance in science. Before the eighteenth century, there was no satisfactory account of what heat and cold were. The best available theory, and here we may use the term to mean "a mere theory," was the suggestion that heat was an extremely light, incompressible fluid that flowed from hotter objects to cooler ones at rates that depended on the density of the objects. Kinetic theory reflects the dawning realization of chemists and physicists that heat was not a separate substance but another manifestation of motion, a phenomenon already well understood since Newton's time in the seventeenth century.

As the nineteenth century progressed, chemists and physicists came to realize that gases are composed of unimaginably vast numbers of particles—molecules of various sizes and masses—which though unobservable may have the same Newtonian properties as observable objects. So the notion arose that heating and cooling gases is a matter of changes in the average values of these Newtonian properties of the molecules gases are made of, as these molecules bounce off one another and off the sides of the gas's container. If a billiard ball can deform ever so slightly the rubber rail of a billiard table, then a hundred million molecules or so hitting the inside of a balloon are likely to do so as well, thus causing it to expand if it is flexible. If the container can't expand because it is rigid, the energy of the molecules must have some other effect. Perhaps, like friction in wheel brakes, which we already know is produced by motion that is resisted, the effect of all these collisions of the molecules with a rigid surface is an increase in heat. And of course if the molecules bounce off each other a great deal, the same increase in heat may arise.

The development of these ideas produced the kinetic theory of gases: (a) gases are made of molecules moving on straight-line paths until they collide with each other or the container, (b) the motion of the molecules—like that of observable objects—is governed by Newton's laws of motion, except that (c) molecules are perfectly elastic, take up no space, and, except when they collide, exert no gravitational or other forces on one another. Given these assumptions, it is relatively easy to derive the ideal gas law from Newton's laws,

$$PV = nRT$$

(where P = pressure on the container walls, V = the volume of the container, r is a constant, and T is temperature in degrees Kelvin). The trick in deriving the ideal gas law is to connect the underlying structure—the behavior of molecules *like* billiard balls—with the measurements we make of the gas's temperature, pressure, and volume. One of the important discoveries of nineteenth-century thermodynamics consists in effecting this connection: showing that the absolute temperature (the quantity of heat) of a gas at equilibrium depends on $1/2\ mv^2$, where m is the mass of an individual molecule and *v* is the average velocity of the ensemble of molecules that constitute the gas in the container. $1/2\ mv^2$ will be recognized in Newtonian mechanics as the mean kinetic energy of all of the molecules.

(We could turn this statement into an identity, if we multiplied the absolute temperature on the right side by 3k/2 where k is the Boltzmann constant, named after one of the important founders of thermodynamics. This constant will make both sides of the equation take on the same units.

$$3k/2\ [\text{T in degrees Kelvin}] = (1/2\ mv^2)$$

Again, $1/2\ mv^2$ is the standard expression for kinetic energy in Newtonian mechanics. Here it is attributed to unobservable molecules that are treated *as though* they were elastic spheres—perfect little billiard balls—that collide.)

The discovery that heat and pressure are the macroscopic reflections of molecular motion meant that physicists were able to explain the gas laws—which had been known since the time of Boyle and Charles (and Newton for that matter), in the seventeenth century. If we set temperature equal to the mean kinetic energy of the molecules of the gas (multiplied by some constant), and pressure equal to the momentum transferred per cm^2 per second to the sizes of the container by the molecules as they bounce off of it, we can derive the ideal gas law (and other laws it subsumes, Boyle's law, Charles's law, Guy Lussac's law) from Newton's laws applied to molecules. We can also derive Graham's law, according to which different gases diffuse out of a container at rates which depend on the ratio of the masses of their molecules, and Dalton's law that the pressure one gas exerts on the walls of a container is unaffected by the pressure any other gas in the container exerts on it. We can even explain Brownian movement—the phenomenon of dust motes in the air remaining in motion above the ground never dropping towards the ground under the force of gravity: they are being pushed in random paths by collision with gas molecules that compose the air. There is in principle no end to the regularities about different types, amounts, and mixtures of particular gases we can derive from and thereby explain by the kinetic theory of gases.

Let's generalize a bit from this case. The kinetic theory of gases consists of Newton's laws of motion plus the law that gases are composed of perfectly elastic point-masses (molecules) which obey Newton's laws, plus the law that the temperature of a gas (in degrees Kelvin) is equal to the mean kinetic energy of

these point-masses, plus some other laws like this one about pressure and volume of the gas.

The kinetic theory thus explains observable phenomena—the instrument-readings we collect when we measure changes in temperature, pressure of a gas, holding volume constant, or pressure and volume, holding temperature constant, etc. The theory does so by making a set of claims about invisible, unobservable, undetectable components of the gas and their equally unobservable properties. It tells us that these components and their properties are governed by laws that we have independently confirmed as applying to observable things like cannon balls, inclined planes, pendula, and, of course, billiard balls. The kinetic theory thus provides an example of one way that the components of a theory work together to explain observations and experiments.

The kinetic theory of gases can illustrate several further components of an approach to the nature of theories that emerged naturally out of the deductive-nomological or covering law approach to explanation that we elaborated in Chapter 3. This approach is usually described nowadays as the axiomatic or **syntactic approach to scientific theories**. It is associated with a view of the way that theories are tested known as "hypothetico-deductivism," according to which scientists theorize—frame hypotheses—but do not test them directly, because like most theories in science they are typically about processes that we cannot directly observe. Rather the scientist deduces testable consequences from these hypotheses. If the tests are borne out by observation, the hypotheses are (indirectly) confirmed. Thus, this axiomatic approach to theories is sometimes called the "hypothetico-deductive" or H-D account of theories.

The axiomatic approach begins with the notion that theories are, as we have suggested, axiomatic systems, in which the explanation of empirical generalizations proceeds by derivation or logical deduction from axioms—laws not derived but assumed in the axiomatic system. Because the axioms—the underived laws fundamental to the theory—usually describe an unobservable underlying mechanism (like our point-mass billiard ball-like gas molecules), they cannot be tested directly by any observation or experiment. These underived axioms are to be treated as hypotheses *indirectly* confirmed by the empirical laws derivable from them, which can be directly tested by experiment and/or observation. It is from these two ideas, that the foundations of a theory are *hypotheses* supported by the consequences *deduced* from them, that the name hypothetico-deductive model derives.

One theory's underived axioms are another theory's explained theorems, of course. Every theory leaves something unexplained—namely those processes that it invokes to do the explaining. But those processes unexplained in one theory will presumably be explained in another. For example, the balanced equations of chemical stoichiometry (for example $2H_2 + O_2 \rightarrow 2H_2O$) are explained by assumptions that the chemist makes about electron-sharing between hydrogen and oxygen atoms. But these laws, underived in chemistry, are the derived, explained generalizations of atomic theory. And atomic theory's assumptions about the behavior of electrons that result in the chemical bond, are themselves

derived in quantum theory from more fundamental generalizations about the components of micro particles. No one suggests that scientists actually present theories as axiomatic systems (though Newton did so), still less that they explicitly seek the derivations of less fundamental laws from more fundamental ones. It is important to remember that like the covering law model, the axiomatic account of theories is a "rational reconstruction" of scientific practice designed to reveal its underlying logic. Nevertheless it claims to have found vindication in both the long-term history of science, and in important theoretical breakthroughs of recent science.

Consider one final case: the accomplishments of Watson and Crick, the molecular biologists who discovered how the chemical structure of the chromosome—the chains of DNA molecules of which it is composed—carry hereditary information about traits from generation to generation. Watson and Crick's theory about the molecular structure of the gene enables geneticists to explain heredity by explaining (in part) the laws of another theory—Mendelian genetics—laws about how hereditary traits, like eye color, are distributed from generation to generation. How did this happen? In principle the situation should be little different from the derivation of the ideal gas law, $PV = nRT$, from the kinetic theory of gases: Given the identification of the gene with a certain amount of DNA, the Mendelian laws governing the segregation and assortment of genes from generation to generation should be logically derivable from a set of laws governing the behavior of DNA molecules. One reason this should be so is of course that a gene is nothing but a strand of DNA—which is what Watson and Crick discovered. So, if Mendel discovered laws about genes, it stands to reason that they obtain in virtue of the operation of laws about DNA molecules. And if this is so, then how more clearly to explain Mendel's laws than by showing that they obtain in virtue of another set of laws, and how more clearly to show this than to logically derive the former from laws of molecular biology.

Alas, matters are not so simple. Mendel's laws, and population genetic theory, are explained in part at least by regularities or laws governing the behavior of DNA, but it turns out that an attempt to derive Mendel's laws from molecular biology is doomed to failure for several reasons (some of which are explored in Chapter 13). This fact makes it mysterious how more fundamental theory in biology explains less fundamental theory, and it even suggests to some philosophers of science (and biologists) that Mendel's laws are not derivative from more fundamental laws to be discovered by molecular biology. In this respect they would be quite different from the ideal gas law that is derivable from more basic theory. These philosophers and biologists suggest that higher-level theory in their discipline is not explainable by derivation or perhaps not explainable at all by lower-level theory. On this basis they argue that biology and the other special sciences are independent of and different from physical theories. This is an important issue with significant metaphysical and methodological implications to which we will turn in the next chapter's discussion of reductionism.

The Philosophical Significance of Newtonian Mechanics and Theories

This process whereby more basic or fundamental theories explain less general ones, improve on them, deal with their exceptions, and unify our scientific knowledge, has seemed to many philosophers of science to characterize the history of science since the time of Newton. In the 250 years from Newton's time to Einstein's, the ability of Newton's theory to explain so much of what was being independently discovered, and to do it by logical derivation of mathematical formulae, had a profound effect on philosophy.

Philosophers were willing to adopt Newton's theory as something close to an intellectual ideal against which other achievements in science should be measured. This established predictive power and mathematical expression as hallmarks of scientific achievement. Philosophers and scientists recognized that Newton's theory was deterministic: given its laws and a description of the positions and momenta (the product of the mass and the velocity) of bodies in a closed system, the past positions of all the bodies in the system were fixed, as were their future ones. If the world was at bottom Newtonian, because every particle of matter obeyed his laws, then the behavior of everything in the world was deterministic, perhaps including human behavior. So much for free will, some philosophers thought. But the cultural significance of Newton's theory was even greater than this. His theory precipitated a revolution in physics, but also in philosophy, in Western thought, and in civilization as a whole.

For some millennia before Newton, it was widely held by scientists and non-scientists alike that the motion of heavenly bodies, the planets and stars, was governed by one set of fixed laws and the motion of things on and near the Earth governed either by no laws or by another set of laws quite different from those governing heavenly motion. This belief reflected an even more fundamental conviction that the realm of the heavens was perfect, unchanging, incorruptible, and entirely different in material composition from the realm of the Earth. Here on Earth things were thought to happen in irregular ways that showed few patterns—things break down and disorder continually threatens to take over, things grow and die. In short, the Earth was supposed to be a far less perfect world than the heavens.

There was another important feature of this dominant pre-Newtonian world-view. The behavior of everything in the world, indeed all motion, even of the simplest inanimate thing, was goal-directed, purposive, driven towards some end or another, and each different kind of thing had a different end, purpose, or goal which reflected its nature or essential properties—the ones that accorded the thing its identity. As the song goes, "fish gotta swim, birds gotta fly" owing to whatever goal it is that striving for leads to. All the sciences were, in the language of the last chapter, thoroughly teleological in their laws, theories, and the explanations these provided.

The connections between this pre-Newtonian scientific world-view and that of the dominant religions before the scientific revolution are obvious. The heavens

differ from the Earth in their perfection and changelessness. And everything under the Sun had a purpose ordained by the designing deity. Once we identified its purpose, we could explain the behavior of things and diagnose their failures.

The accomplishments of Kepler and Galileo weakened this world-view fatally at the beginning of the seventeenth century. Employing data gathered by Tycho Brahe, the sixteenth-century Danish astronomer, Kepler showed that we could predict the position of the planets in the night sky by assuming that they travel around the Sun on ellipses and that their velocity is a specific mathematical function of their distance from the Sun. Since we are "aboard" one of these planets its actual motion and the motion of the other planets around the Sun is hidden from us. Moreover, the data Brahe collected about the apparent position of the planets in the night sky confirmed Kepler's hypothesis about elliptical orbits. The mathematical exactness of Kepler's laws may have lent some credibility to the ancient notion of the perfection of heavenly processes. But these notions were almost fatally undermined by Galileo's telescopic observations of the Moon, its craters and other irregularities as well as previously undetected sunspots. The imperfections of the latter and the recognizable similarity of the former to terrestrial features couldn't be reconciled with the notion that there is a fundamental difference between perfection and changelessness in the heavens and the generation and corruption of earthly processes.

Galileo's experiments, dropping cannon balls (according to legend) from the Leaning Tower of Pisa, rolling them down inclined planes, timing the period of pendula as their lengths are changed, all contributed to his discovery of the laws of motion of objects in the immediate vicinity of the Earth: projectiles always follow the paths of parabolas, the period of a pendulum (the time for one cycle of back and forth motion) depends on the length of the wire and never the weight of the bob, free-falling bodies of any mass have constant acceleration. The mathematical expression of these laws encouraged the physicists of the sixteenth and seventeenth centuries such as Descartes to seek mathematically formulated explanations of natural processes and to reject teleological ones as empty and without any quantitative applicability in prediction or technology. All this set the stage for Newton finally to change the natural scientist's conception of reality utterly. It did so in two ways: First, it made science's departure from the common-sense view of the world both final and deep. Second, it overthrew the teleological view of the physical universe that had been orthodox since even before Aristotle in fourth-century BC Greece, and confronted teleology everywhere else in science and ordinary life with a mortal challenge.

Newton's great scientific achievement was to show that Kepler's laws of planetary motion and Galileo's laws of terrestrial motion, along with a lot of other generalizations about straight-line and curved motion, pendula, incline planes and buoyancy, can be derived from the single set of four laws given above, which are silent on goals, ends, purposes, essences or natures; these laws mention only completely "inert" brute physical properties of things, their mass, their velocity, their acceleration, and their distance from one another, along with their

gravitational attraction. Newton had unified and derived the greatest scientific explanation of all time.

For all its simplicity, Newton's first law makes a radical break with pre-Newtonian science and with common sense, so much so that many people who know the law still don't realize its significance. As noted above, the first law tells us that whether something is at rest is not a matter of its being in motion or not. Things moving at any speed whatever are at rest just so long as their speed is not changing. Things are at rest, Newton's theory tells us, when they are neither accelerating nor decelerating. This immediately changed the entire explanatory agenda of science. It was no longer necessary to explain straight-line motion at constant speed. There is nothing here that needs explaining. This shift is as profound as it is invisible.

What is more there seems no way to treat the best prior scientific theory—impetus theory—as a stepping stone to the theory that Newton and Galileo before him were formulating. The prior impetus theory, that things in motion are subject to or contain some force, isn't even approximately correct if Newtonian mechanics is right. This complete break with previous theory is a hallmark of what made the seventeenth century the age of *the* scientific revolution. Pre-Newtonian physics (and still today the "folk physics" that most people accept) are in agreement both that when something is in motion, it is not at rest, and that for something to be in motion requires a force to act upon it. That's what impetus theory tells us. This is just what Newton's second law denies: the force on a body is equal to its acceleration times its mass, $F = ma$. When velocity is constant, no matter how high, acceleration is zero and so by Newton's second law, the force acting on the body must be zero as well. Bodies on which no forces act are at rest (i.e. have zero acceleration). And if they have non-zero velocity, moving with zero acceleration, they move in straight lines. Thus, when a body travels along a curved path, it must be the case, according to Newton's laws, that forces are acting upon it, i.e. that its motion in at least one direction is slowing down or speeding up.

Newton's third law is the one that people seem to know best, and seems most intuitive: it is often expressed as the statement that for every action there is an equal and opposite reaction. "Action" is of course a deceptive term in this expression, and probably the source of the conviction that the third law expresses some insight accessible to common sense independent of physics. In the context of Newtonian mechanics, actions are changes of velocity, i.e. events that reflect the "action" of forces on bodies. Things in motion have momentum, which we can feel as they bump into us. Momentum is defined in physics as the product of mass and velocity. Newton employed the third law to derive the conservation of momentum: the total amount of momentum that a set of bodies has remains the same as they bounce into each other. Each body transfers some or all of its momentum to the bodies with which it collides. Since it loses or gains momentum with each collision, if a body's mass remains constant (it doesn't crumble or break apart), its velocity must change. If a set of bodies keep colliding without breaking up (or otherwise losing matter), then the third law says that when you

add up the momenta, or just the velocities at any two times, the totals will have to remain constant.

Of course when we apply all three of these laws at or near the surface of the Earth at normal atmospheric temperature and pressure, to things like soccer balls or feathers, we have to allow for the interference of air molecules, the friction of the ground against a ball, or other conditions, each extremely slight, but, added together, enough to make illustrating Newton's laws difficult. Even a hockey puck on the smoothest iced-over lake will eventually stop. This does not show that Newton's first law is false; it shows that forces are acting on the puck even though they are undetectable to us. In this case friction as the molecular motion of the molecules in the puck heats up the ice and melts it, thereby slowing down the puck (try freezing the puck and see if goes farther).

The one law that Newton formulated which is visibly exemplified to very great precision (by the Moon and the Earth, the planets and the Sun, two binary stars, etc.) is the inverse square law of gravitational attraction. As noted above, the inverse square law is in an important respect different from the other three. Newton's first three laws seem to operate through spatial contact between bodies. If a body is at rest you have to push it or pull it to change its velocity; pushing and pulling are the ways force is applied to things. To get a body that is accelerating in a straight line to speed up further, or slow down, or to change its direction, you have to interfere with its momentum, again by applying a force, pushing or pulling it away from its original path. These laws seem to reflect the corpuscularianism that went hand in hand with the scientific revolution's denial of teleology or purpose in physics. For this reason corpuscularian scientists and philosophers (who were one and the same in the seventeenth century) sought a corpuscularian theory of gravitational phenomena. The most famous of these was Descartes' "vortex" theory, which held that the apparently empty space between the Sun and the planets is filled with corpuscles spinning and thereby communicating forces that move planets in their orbits around the Sun. Newton advanced a number of objections to this theory. But he sought corpuscularian approaches to gravity before finally biting the bullet and admitting the existence of a force that by his own philosophy should have been treated as "occult." As noted above, gravity acts "at a distance" even through the empty space of a vacuum, so no corpuscles can be involved. It is "felt" everywhere so it must move at infinite speed, which, given Newton's other laws, would require infinite momentum if it had any mass at all, as corpuscularianism requires of what is physically real. Finally, it is difficult to understand how a force carried along by any corpuscles can penetrate any barrier whatever and again with infinite speed. So, in every respect gravity is a problematical concept for Newton. Gravity should have been a major metaphysical and epistemological embarrassment, to physics and to science.

But far from undermining corpuscularianism, gravity was simply compartmentalized by Newton and the corpuscularians, and ignored as a problem that could not be solved but some day might be. In response to the obvious tension between his corpuscularianism and his inverse square law

Newton famously asserted, "*Hypotheses non fingo*," meaning roughly, "In the absence of experimental data about the physical basis of the behavior described by the inverse square law, I refuse to speculate." This combination of a stern injunction against doing science in the absence of experiment, and a refusal to acknowledge an inconsistency, successfully protected physics from prematurely and unproductively dealing with the problem of the nature of gravity until Albert Einstein got around to it in 1912 or so.

Meanwhile the explanatory and predictive power of Newtonian mechanics gave it great epistemological and metaphysical significance. Newton's four laws became the paradigm case of what a scientific law should be: universal in form, apparently unencumbered by *ceteris paribus* clauses (until electrostatic forces were discovered), and continually applicable to wider and wider ranges of phenomena. They were so formidable that, as we have seen, Kant felt he had to explain why Newton's laws were necessarily true. Even those who held Newton's laws to be contingent viewed them as fundamental truths about the universe. They also held that the properties that figure in these laws, mass, velocity, acceleration, force, and the properties like energy, and momentum that can be defined in terms of them, are the fundamental features of the world, nature or reality, and that all other properties of things would eventually be explained as constituted by corpuscles with different amounts of these properties in different arrangements. For these scientists and philosophers the great challenge was to discover how Newtonian processes could produce, first, biological phenomena and, eventually, human thought and behavior. For those who opposed the mechanistic world-view extrapolated from Newton's laws, the challenge was to show the opposite. They needed to show that no such extension of the theory was possible and that the biological realm was not merely a subdivision of physics, made of merely more complicated but ultimately purely physical matter in motion.

Additionally, Newtonian mechanics gave strong grounds for determinism about everything in nature. Given only the position and momentum of a body, no matter how large or small, Newton's laws governed its entire future and past trajectories through time and space. If the laws hold for all physical objects, including all living things (including us), then our behavior must be determined as strictly as that of the planets in their paths around the Sun. Insofar as everything that happens to us consists in the interaction of corpuscles within our bodies, all of our behavior is determined as well. Proponents of human free will had to wrestle with this argument. Proponents of the explanatory power of Newtonian mechanics either had to deny that there is any free will or seek ways to reconcile its existence with the determinism that Newton's theory committed them to. There are many possible avenues of argument against the claim that the truth of Newton's laws deprives human beings of free will. But each of them requires a great deal of philosophical ingenuity. These problems are not really ones that the philosophy of science faces, but their importance in intellectual life is testimony to the cultural importance of Newtonian mechanics.

Summary

Theories are sets of laws that work together to explain empirical regularities by derivation of those regularities, and often also by explaining the exceptions and counterexamples that those regularities face. But making precise how the component laws of a theory do this is not easy and almost certainly requires us to come to grips with issues of causality and with the need for claims about unobservable phenomena. These unavoidable topics are ones that are difficult for empiricists to deal with. Since empiricism is the "default" or official epistemology of the sciences, the nature of theories and their central role in all science raises a series of difficulties for the philosophy of science.

Meanwhile, independent of the epistemological problems that the nature of theories raise for philosophy, it is important to emphasize the broader conceptual and historical importance of some scientific theories, especially Newtonian mechanics. Its achievement in explanation and the systematization of knowledge throughout physics over many centuries has revolutionized the landscape of Western thought. This has made it worth working through the leading ideas of Newtonian mechanics and sketching how they overturned a world-view that had ruled science and civilization for at least 2,000 years.

Study Questions

1. Is Euclidean geometry a scientific theory or a mathematical system?
2. Should all theories be expressed as axiomatic systems, the way that Newton's *Principia* did? What advantages and disadvantages might this have?
3. Aristotelian impetus theory was predictively accurate about many terrestrial phenomena. How could this be if it was based on a complete mistake about the nature of rest and forces?
4. Why is Newton's fourth law, the inverse square law, not a theorem derivable from the other three laws? What does it add to the description of nature that they are silent about?
5. Why should anyone suppose a theory about inanimate objects, like Newton's, has any implications for living things, like animals and people?
6. Why was Newton uneasy about his greatest scientific discovery and theoretical innovation, gravity?

Suggested Readings

The history of philosophical analysis of scientific theorizing is reported in F. Suppe, *The Structure of Scientific Theories*. The axiomatic approach was perhaps first fully articulated in R. Braithwaite, *Scientific Explanation*. Perhaps the most influential and extensive account of theories, and of science in general, to emerge from the period of logical empiricism is E. Nagel, *The Structure of Science*, first published in 1961. This magisterial work is worthy of careful study on all topics in the philosophy of science. Its account of the nature of theories, its development of examples, and its identification of philosophical issues remains unrivaled. Nagel's discussion of the structure of theories, of reductionism and the realism/antirealism issue set the agenda for the next several decades. Two extracts from this work are to be found in Balashov and Rosenberg, "Experimental Laws and Theory," which discusses the relationship between theories and the generalizations they explain.

Thomas Kuhn's *The Copernican Revolution* is an excellent introduction to the details of the scientific revolution initiated by Brahe, Kepler, Copernicus, and Galileo. Steven Shapin's *The Scientific Revolution* is a good introduction to the history of seventeenth- and eighteenth-century science. *Never at Rest* by Richard Westfall is a detailed scientific biography of Isaac Newton. Andrew Janiak treats Newton's philosophy of science and its relationship to his science in *Newton as Philosopher*. Janiak, *Newton: Philosophical Writings*, is a convenient source for Newton's own thoughts about scientific method.

Two classic texts on the philosophical implications of the work of Newton and others in the scientific revolution of the seventeenth century are E. A. Burtt, *The Metaphysical Foundations of Modern Science*, and A. N. Whitehead, *Science and the Modern World*. There are more suggested readings on this subject at the end of Chapter 12.

8 Epistemic and Metaphysical Issues about Scientific Theories

Overview

Theories are indispensable to scientific understanding, to the unification of scientific knowledge, to deepening its explanations, to increasing the precision of its predictions, and to furthering its technological applications. In addition, as we saw in the last chapter, some theories are so wide in their implications that they effect revolutions in science and in the culture more broadly.

But at the same time there are features of scientific theories that raise profound questions about the nature, extent, and justification for the claims to knowledge about the world based on them. To the extent that these epistemological questions are unaddressed or unanswered, their grounds and implications may be disputed. In fact, it is in part because of the problems they raise for human knowledge that scientific theories remain controversial even among scientists, some philosophers, and many ordinary people.

In this chapter and the next we begin exploration of these epistemic issues. They will arise again throughout the rest of this book. First we consider the nature of scientific progress, which is traditionally portrayed as revealing how nature is made sense of via the unification of scientific theories. But the way in which theories have been unified in physical science in particular raises difficulties for empiricism in accepting scientific theories at all. If empiricism is incompatible with theorizing, then presumably the problem lies with the former? Few philosophers of science are prepared

to reject scientific theories based solely on epistemological worries. On the other hand, scientists insist on deciding among theories empirically. These approaches to the problem of grounding theory in evidence take us from epistemology to unavoidable questions about the metaphysics of science—whether we should treat its existence claims as true or even approximately true.

Reduction, Replacement, and the Progress of Science

In showing that Kepler's and Galileo's laws were but special cases of more general laws that were true everywhere and always, Newton not only explained why these laws obtained, but also undercut the basic metaphysical conviction that the realm of the heavens was somehow different from that of Earth. As Chapter 7 sketched, along with Galileo's telescopic discovery of the craters and other imperfections of the Moon and the Sun, Newton's revolution had a profound intellectual influence far beyond the formal derivation that he provided to unify physical theory. The power of Newton's unification was further increased in the ensuing 200 years as more and more phenomena came to be explained (or explained in more quantitative detail) by it. Eclipses, the period of Halley's comet, the spheroid shape of the Earth, the tides, the precession of the equinoxes, buoyancy and aerodynamics, parts of thermodynamics, all were unified and shown to be part of "the same underlying process," through the derivation of laws describing these phenomena drawn from Newton's four fundamental laws.

None of these laws, moreover, appealed to future goals, purposes, or ends. Instead, all identified prior or present causes (position and momentum), and all except the inverse square law identified forces that act through physical contact as sufficient to explain physical processes. As such, Newtonian mechanics allowed us to dispense completely with purposes, goals, and ends as properties that pre-Newtonian science invoked to explain the behavior of physical systems. The success of Newtonian mechanics thus encouraged a world-view according to which the physical universe was just a vast "clockwork" mechanism in which there was no teleology of the sort we discussed in Chapter 6.

Of course Newton's theory could not explain the behavior of living things, though some "mechanists" among scientists and philosophers held out the hope that it would eventually explain everything in terms of deterministic laws about position, momentum, and gravity. Biology, however, remained a safe haven for teleological explanations long after it was eliminated from physical science. Kant, who as we saw in Chapter 2 argued that Newtonian mechanics was necessarily true of the physical world, held that its purely mechanistic picture of the physical world could never be extended to explain the biological realm. Recall Kant's statement that there will "never be a Newton for the blade of grass." As

with his claims about the necessity of Newton's laws, this claim of Kant's was also overtaken by events.

Newton showed how Galileo's and Kepler's laws could be derived from his own theories as special cases. Philosophers of science refer to this derivation of the laws of one theory from the laws of another as "intertheoretical reduction" or simply "**reduction**." Reduction requires that the laws of the reduced theory be derived from that of the reducing theory. If explanation is a form of derivation, then the reduction of one theory to another explains the reduced theory; in effect it shows that the axioms of the less basic theory are theorems of the more basic one.

So the scientific revolution of the seventeenth century appears to consist in the discovery and reduction of Galileo's and Kepler's laws to Newton's. What is more, the progress of physics from the sixteenth century onwards seemed to be the history of less general theories being successively reduced to more general ones, until the twentieth century, when suddenly theories even more general than Newton's were framed, which in turn reduce Newtonian mechanics by derivation: the special and general theories of relativity and quantum mechanics. Newton's laws are deducible from the laws of these theories by making some idealizing assumptions, in particular that the speed of light is infinite or at least that all other attainable velocities are much, much slower than the speed of light, and that energy comes in continuous amounts and not in discrete but very small units or "quanta."

One simple example often used to show that Newton's theory turns out to be a special case of Einstein's theory of special relativity employs the well-known "Lorenz contraction" equation:

$$\text{Length}_{\text{measured by observer in motion}} = \text{Length}_{\text{at rest}} \sqrt{1 - {}^{v^2}/_{c^2}}$$

The equation expresses the special theory of relativity's consequence that the length of an object as measured by an observer in motion with respect to that object is shorter than its length measured at rest with respect to the object. This difference is given by a factor of the square root of 1 minus the velocity squared divided by the speed of light squared. This number is usually very close to 1, so the length contraction is not detectable. Newton's theory requires that lengths not vary with the motion of the measurer. A similar equation governs relativistic mass and Newtonian, or rest mass. We can derive Newton's laws from Einstein's theory as a special case where the measurer's velocity is very small compared to the speed of light, which it almost always is. In this special case, relativistic versions of the laws of motion reduce to versions of Newton's theory.

According to one traditional view in the philosophy of science, the reduction of less fundamental theories to more fundamental ones reflects the fact that science is successively enlarging its range and depth of explanation as more and more initially isolated theories are shown to be special cases, derived from

a smaller and smaller number of more fundamental theories. Scientific change is scientific progress and progress comes in large measure through reduction. In fact, reduction is also viewed as the characteristic relation among disciplines once they attain the status of sciences.

This view of the relations between theories and disciplines in science goes under the label of the "unity of science." Sometimes this thesis is understood to be the more limited epistemological claim that the sciences all share the same method of knowledge-acquisition. Broadly speaking, they all adopt a commitment to the control of inquiry by observation, experiment, data collection, and other uncontroversially empirical means. A stronger version of this thesis, however, adds a metaphysical component: That the reality that the sciences explore is also unified—that it consists only in physical things and processes, and combinations and aggregations of them. This stronger version of the unity of science thesis is the one usually understood and widely disputed.

On this view physics is the most basic science since its domain is the behavior of all things and it explains them in terms of the fundamental constituents of matter and fields. Chemistry is next as it explains everything more complex than individual atoms. Biology explains organic molecules and their aggregation, and so on in a hierarchy of increasingly less basic sciences. The sciences in this hierarchy are unified by derivation or reduction. Thus, in principle, chemistry should be reducible to physics via atomic theory, and biology should be reducible to chemistry via molecular biology. Similarly, we should seek a psychological science composed of laws that are themselves reducible to the laws of neuroscience—a division of biology. Of course the social sciences have yet to uncover (and may never uncover) laws reducible to those of natural science, via reduction to psychological laws. Therefore, it is argued by proponents of the unity of science thesis, these disciplines lack an important feature common to scientific theories—linkage via reduction to the most fundamental and predictively powerful of the sciences, physics. If the human sciences or some theories in them are not reducible to one or another of the established theories in the natural sciences, then on the thesis of the unity of science, they must be guilty of methodological errors, and cannot yet be counted as scientific theories or disciplines. The unity of science thesis was held at least by some logical positivists and their followers. It appeared to be vindicated by developments in the history of science from Newton's time to the twentieth century. Important physicists still endorse this view or ones that approach it.

The unity of science thesis makes axiomatization attractive as an account of how a theory explains by uncovering more general underlying mechanisms that systematize and explain less general ones. If the universe reflects the neat picture of layers of causal laws, each of which rests on a layer of more fundamental laws below it (that logically imply the less fundamental laws) and if the universe is composed of a small number of basic kinds of things that behave in a uniform way (out of which everything else is composed), then there should be a uniquely correct description of nature. The description takes an axiomatic form because reality is a matter of the complex being built up out of the simple in accordance

with general laws working together. The commitment to axiomatization as giving the structure of theory and the relations among theories supports a metaphysical claim about the nature of reality: at bottom it is simple in composition and operation, and all the complexity and diversity of more complicated and more composite things is the result of their material dependence on the most basic structures.

Of course, this picture must be substantially complicated. To begin with, the notion that the laws of one theory may be directly derivable from those of another is too simple. Scientific progress involves the correction and improvement of a theory's predictions and explanations by its successors. If the successor theory merely "contained" the original reduced theory as a logical consequence, it would incorporate the errors of its predecessor. For example, Galileo's law of terrestrial motion implies that the acceleration of bodies falling towards the Earth remains constant, while Newton's laws recognize that accelerations must increase owing to the gravitational force between the Earth and bodies approaching it. For predictive purposes we can neglect these slight increases in acceleration, but we must correct Galileo's terrestrial mechanics, by adding gravitational force, if it is to follow from Newton's laws. Similarly, Mendel's laws of genetics should not follow directly from laws in contemporary molecular genetics, for we know that Mendel's laws are wrong. Phenomena like genetic linkage and gene cross-over falsify these laws. What we want in any reduction of Mendel's laws to more fundamental laws of molecular genetics is therefore an explanation of the exceptions, where Mendel's laws go wrong as well as where they work. This suggests that reduction usually involves deriving a "corrected" version of the theory to be reduced from the more fundamental reducing theory.

But the requirement that the reduced theory must sometimes be "corrected" creates problems for the axiomatic view of theory change. Sometimes, one theory supersedes another not by reducing it, but by replacing it. As noted in Chapter 7, the shift from treating rest as zero velocity in Aristotelian physics to treating it as zero acceleration in Newton's theory reflects such a replacement. Indeed, replacement seems characteristic of a discipline's becoming a "real" science. For example, before the work of Lavoisier in the late eighteenth century, combustion was explained by "phlogiston" theory. Phlogiston was hypothesized to be a substance that escapes from things when they burn, but owing to its character could not be directly observed. One trouble with phlogiston theory is that careful measurements revealed that burning a substance increases its weight. Therefore if phlogiston is liberated in combustion, it must have negative weight. Since weight depends on mass and on the strength of the Earth's gravitational force, which presumably remains constant when things burn, it would seem that phlogiston has negative mass. This is something hard to reconcile with Newtonian physics. For this and other reasons, chemists were dissatisfied with phlogiston theory despite some of its apparently satisfactory explanations of chemical experiments in combustion. Lavoisier advanced a new theory, which hypothesized a quite different, unobservable substance, which he termed "oxygen," which is incorporated by substances when they burn and so, among other things, need not have negative mass.

Lavoisier's oxygen theory therefore did not reduce the older phlogiston theory of combustion. It *replaced* the "ontology"—the kinds of things phlogiston theory was about: phlogiston, dephlogisticated air, etc.—and its alleged laws. Lavoisier's theory provided for a completely different kind of thing, oxygen, which could not be linked up to phlogiston in ways that would enable this latter concept to survive. Attempts to define phlogiston in terms of the concepts of Lavoisier's theory of combustion will not enable us to derive the phlogiston theory from Lavoisier's theory. And of course, Lavoisier's theory is the beginning of modern chemistry. Accordingly, scientists say that there never was any such thing as phlogiston. This break between theories that cannot be linked reductively or otherwise is like the one between Aristotelian impetus theory and Newtonian inertial theory. Lavoisier's achievement of establishing chemistry as a science is confirmed by the reduction of oxygen theory to atomic physics 150 years later.

By contrast to the case of phlogiston and oxygen, when a theory is reduced to a broader or more fundamental one, the "ontology" of the reduced theory—the kinds of things it makes claims about—is preserved. The reason is obvious. If reduction is a matter of deduction of the law of the reduced theory from those of the reducing theory, such derivation is possible only when the terms of the two theories are connected. To derive the laws of Mendelian genetics from those of molecular genetics you must first define the Mendelian gene in terms of nucleic acids. For it is assemblages of DNA that molecular genetics are about and Mendelian genes that Mendel's laws are about.

In general, a law about all As being Fs will only follow from a law about all As being Bs if every B is identical to a C and every C is identical to an F. Indeed a large measure of the achievement of reduction is the formulation of these identities. For example, the reduction of the thermodynamics of gases to statistical mechanics turns on the identity we noted in Chapter 7:

$$3k/2 \text{ [T in degrees Kelvin]} = (1/2 \ mv^2)$$

Whether we treat this identity as a definition or a general law relating temperature to kinetic energy, its formulation was the crucial breakthrough that enabled physicists to reduce the behavior of gases to the behavior of the molecules that compose them.

The trouble for reduction as an account of scientific progress across history from Newton's day to our own is that such intertheoretical identifications became increasingly difficult. Defining the properties that figure in thermodynamics in the vocabulary of Newtonian mechanics was a great achievement, but obviously a feasible one. As physics developed, however, such unifying identifications became more difficult to establish. For example, the second law of thermodynamics tells us that entropy will probably increase. But no one has succeeded in defining entropy in terms of the more fundamental concepts of Newtonian mechanics—mass, velocity, and force. Yet no one is prepared to dispense with entropy the way that phlogiston and impetus were cast aside as science progressed.

The situation is even more serious in biology. Here we find a host of concepts—gene, chromosome, nucleus, organelle, cell, tissue, organ, organism—none of which can be defined in terms of their chemical constituents, still less their physical ones—atoms, electrons, protons, quarks, etc. As noted in Chapter 7, Watson and Crick are supposed to have discovered what genes are made of—DNA—and their discovery immediately suggested to them how genes transmitted hereditary information in the way that Mendel's laws require. But no one has ever been able to derive Mendel's laws from regularities in molecular biology, and no one seems to feel the need to do so, either. The reason it cannot be done is that despite Watson and Crick's discovery, the Mendelian gene cannot be completely defined in terms of DNA sequences, not even in principle. This is owing to the complexity, redundancy, and disparity of DNA structures that constitute any single gene or the whole class of them. So, it is difficult to see how laws in genetics can be derived from those of chemistry or physics. Yet no one is prepared to dispense with these laws or other biological theories as unscientific. (For more on Mendel's laws and their reduction, see Chapter 13.)

After the eclipse of logical positivism, reductionism became an unpopular doctrine in the philosophy of science. For one thing it is evident that in many disciplines, especially the life sciences and the behavioral sciences, there are no exceptionless laws or universal regularities to be found. In these disciplines theories were constituted by models, each of which describe a much more limited range of cases. There was nothing in these sciences to be "derived" from the laws of more basic sciences. As the study of the life sciences especially began to attract the attention of philosophers of science away from their traditional focus on physics, this problem for reductionism became acute. By the beginning of the twenty-first century, using the term 'reductionism' as the label for a methodology or research strategy was supplanted by the term 'mechanism' to describe the demand that science, especially biology, uncover mechanisms (ultimately molecular, physical mechanisms) and that models of biological systems and processes be explained by identifying the chemical and physical mechanisms that compose what they model.

Exponents of mechanism explicitly recognized that especially in the life and social sciences, the goal was to construct limited models of specific processes, not general theories composed of universal laws. Accordingly, they argued that *successful* explanatory models in these sciences had to show that the variables in any explanatory model 'correspond' to components and activities of component parts that make them up, and causally bring about the processes described in the models. Mechanism thus demands a reduction of wholes to parts, instead of laws to more fundamental laws. Mechanists held that without providing such details about underlying mechanisms, even a predictively successful mathematical model—a set of equations correctly relating dependent and independent variables—could not explain the phenomena it enabled the scientist accurately to predict. In effect mechanists adopted a strong causal account of explanation and insisted that causation is a matter of the interaction of component parts of a system working together in a mechanism that the scientist is required to reveal.

Mechanists thus view the domain of the life sciences not as a hierarchy of theories deductively organized, but as a series of nested "black" boxes, each to be made transparent by being opened up, with its workings explained by having its parts identified. These parts are then in turn the next set of black boxes to be illuminated.

As Chapters 9 and 10 will show, just as with reductionism, the most serious qualm about mechanism is its implication for the autonomy of higher-level explanations. Mechanism shares with reductionism the requirement that explanation in the life sciences and the special sciences be reductive—that higher-level models, especially ones involving functional capacities, be "cashed in" for lower-level component mechanisms plus the organization of them that provides a causal process realizing the model. It thus threatens the view that there are *autonomous* causal explanations in biology, that the explanations of higher-level science illuminate phenomena in the domains of their disciplines even without the identification of mechanism that "lower-level" sciences might provide.

The upshot is serious: Either the history of science is progressive and it proceeds by reduction of theories (but we need a very different account of reduction that does not treat it as theoretical explanation by logical derivation) *or* the history of science is progressive, but this progress is not a matter of reduction at all. Rather it is a matter of the replacement by increasingly correct but radically different theories. Or, more radically, one might conclude that despite appearances the history of science is not progressive at all, but rather reflects some other much more complicated succession of theories. Few philosophers of science ever considered the third alternative until the 1970s. We will devote considerable space to this view in our discussion of the work of Thomas Kuhn in Chapter 12. Meanwhile, the problem of defining the characteristic concepts of less fundamental theories in terms of the concepts of more fundamental ones raises an even more troubling issue: how do any of these concepts, especially the most fundamental ones, acquire their meaning in the first place?

Reductionism makes the relationship between evidence and explanation mysterious. It seems a characteristic feature of reduction that it unifies observable phenomena or at least unifies the generalizations that report them to more and more fundamental, more and more accurate, regularities that are more and more observationally inaccessible. Having begun with cannon balls and planets, physics finally succeeds in explaining everything in terms of undetectable microparticles and their properties. So, it seems to make explanatorily basic the very thing that is epistemically most problematic. While the official epistemology of science is empiricism—the thesis that our knowledge is justified only by experience, that is, experiment and observation—its explanatory function is fulfilled by just those sorts of things that creatures like us cannot experience directly. Indeed, the microparticles of modern high-energy physics are things that no creature like us could possibly have any acquaintance with. And this fact raises the most vexing questions about the nature of scientific theories.

The Problem of Theoretical Terms

Scientific explanations are supposed to be testable, they have "empirical content," their component laws describe the way things are in the world, and they have implications for our experience. But almost from the outset science has explained by appeal to a realm of unobservable, undetectable, theoretical entities, processes, things, events, and properties. As far back as Newton, physicists and philosophers have been uncomfortable about the fact that such things seem both explanatorily necessary and unknowable. Unknowable, because unobservable; necessary because without appeal to them, theory cannot effect the broad unification of observations that the most powerful explanations consist in. Gravity is a good example of the problem.

Newtonian mechanics makes sense out of a vast range of physical processes by showing how they are the result of contact between bodies with mass. We can explain the behavior of a wind-up clock, for example, by tracing a causal chain of cogs, wheels, weights, hour and minute hands, chimes, and twittering birds in which the pushes and pulls that our observations detect are quantified and systematized into exchanges of momentum and conservation of energy between things that are in contact with one another. And this mechanical explanation will itself presumably give way to an even more basic explanation in terms of the mechanical properties of the component parts of the cogs and wheels, and in turn the mechanical properties of their parts until at last we have explained the behavior of our clock in terms of the behavior of the molecules and atoms that compose it. This at any rate is the explanatory expectation of the reductionist.

By contrast, as we have seen, Newtonian gravity is not a "contact" force. It is a force that is transmitted across all distances at infinite speed apparently without any energy being expended. It moves continually through total vacuums, in which there is nothing to carry it from point to point. Unlike anything else it is a force against which nothing can shield us. And yet it is a force itself completely undetectable except through its effects as we carry masses from areas of greater gravitational force (like the Earth) to areas of lesser gravitational force (like the Moon). All in all, gravity is a theoretical entity so different from anything else we encounter in our observations, that these observations do not help us much to understand what it could be. Gravity is a thing so different from other causal variables that one might be pardoned for doubting its existence, or at least being uncomfortable in invoking it to explain anything. One would not be surprised by a centuries-long search for some "mechanical" explanation of how gravity works or even better some less mysterious substitute for it.

Most of Newton's contemporaries felt this discomfort with the notion of gravity. But neither they nor later physicists were prepared to dispense with the notion. For dispensing with gravity means giving up the inverse square law of gravitational attraction,

$$F = gm_1m_2 / d^2$$

and no one is prepared to do this. Gravity thus seems an "occult" force, whose operation is no less mysterious than those that non-scientific explanations like astrological horoscopes invoke to allay our curiosity. And the same may be said of other such unobservable notions. Thus, the molecules that compose a gas are supposed to have the properties of little billiard balls, for it is their billiard ball-like behavior that explains the ideal gas law. But if gas molecules are small masses, then surely they are colored, for nothing can be a mass unless it takes up space, and nothing can take up space unless it has some color. But individual molecules have no color. So, in what sense could they be small masses? The obvious answer is that unobservable things aren't just small versions of observable things; they have their own distinct properties—charge, quantized angular momentum, magnetic moments, etc. But how do we know this if our knowledge is justified only by what we can experience through our senses? And, as noted above, by what right can we claim that theories invoking these theoretical entities and properties provide real explanations when we cannot have experience of them? Why should a theory about electrons or genes that we cannot see, touch, smell, taste, or feel be any better at explanation than astrology, New Age mystery-mongering, superstition or fairy-tales?

We can express our problem of justification as one about the meaning of words and the learnability of language. Consider the terms we employ to describe our experiences: the names for observable properties of things—their colors, shapes, textures, smells, tastes, sounds. These terms we understand because they name our experiences. Then there are the terms that describe objects that have these properties—tables and chairs, clouds and clocks, lakes and trees, dogs and cats, etc. We can agree on the meaning of these terms, too. Furthermore, it is tempting to suppose that all the rest of our language is somehow built up out of the names for sensory properties and the labels for everyday objects. For otherwise, how could we have ever learned language? Unless some words are defined not by appeal to other words, but by the fact that they label things that we can directly experience, we could never learn any language. Without such extra-linguistically defined terms we could not break out of a never-ending circle or regress of definitions of one word by reference to other words, and those words defined by reference to still other words, and so on. We would already have to know a language in order to learn it.

Furthermore, language is an infinite disposition: we can produce and understand any of an indefinite number of different sentences. Yet we can do so on the basis of a finite brain that has learned to speak in a finite amount of time; it is hard to see how we managed this feat unless language is either somehow innate or there is some basic vocabulary from which all the rest of language is built up. Now the hypothesis that language is innate (as opposed to a language-learning device) is one that empiricists and most scientists have never taken very seriously. We were not born knowing any language; otherwise it would be hard to see how it is that any human child can with equal facility learn any human language, from birth. That leaves the hypothesis that we learned a finite stock of basic words of one language, which together with composition rules enables us

to build up the capacity to produce and understand any of an infinite number of sentences of that language. What else could this finite stock be but the basic vocabulary we learned as infants? And this vocabulary is of course the names of sensory experiences—hot, cold, sweet, red, smooth, soft, etc., along with words like mom and dad.

But if this is the basis of language, then every word with a meaning in our language must ultimately have a definition in terms of words that name sensory properties and everyday objects. And this requirement should include the theoretical terms of modern science. If these words have meaning, then they must somehow be definable by appeal to the fundamental vocabulary of human experience. This argument goes back to eighteenth-century British empiricist philosophers like Berkeley and Hume. These philosophers were troubled by the "secret powers" like "gravity" and unobservable things like "corpuscles" invoked in seventeenth-century physics. The word "gravity" must lack meaning since it names no experience that people have. The empiricists' disquiet about these theoretical entities has had a continuing impact on the philosophy of science right up to the end of the twentieth century and beyond it.

As we saw in Chapter 2, the twentieth-century followers of the British empiricists labeled themselves positivists and logical empiricists. The logical empiricists inferred from arguments about the learnability of language that the theoretical vocabulary of science had ultimately to be "cashed in" for claims about what we can observe, on pain of otherwise being simply empty, meaningless noises and inscriptions. These philosophers went further and argued that much of what in the nineteenth and twentieth centuries passed for scientific theorizing could be shown to be meaningless nonsense, because its theoretical terms were not translatable into the terms of ordinary sensory experience. Thus, Hegel's physics (quoted from in Chapter 2), Marx's dialectical materialism, and Freud's psychodynamic theory were stigmatized as pseudoscience, because their explanatory concepts—surplus value, the Oedipus complex, etc.—could not be given empirical meaning. Similarly a whole host of biological theories that postulated "vital forces" were denied explanatory power by these philosophers because they invoked entities, processes, forces which could not be defined by appeal to observations. But it was not just pseudoscience that these empiricist philosophers attacked. As we have seen, even such indispensable terms as "gravity" were subject to criticism for lack of "empirical content." Some logical positivists, and the nineteenth-century physicists who influenced them, also denied the meaningfulness of concepts such as "molecule" and "atom." For such empiricists a concept, term, or word has empirical content only if it names something or some property we could have sensory awareness of.

Of course, empiricists held that there would be no problem invoking theoretical entities if the terms we used to name them could be defined by way of observable things and their properties. For in that case not only would we be able to understand the meaning of theoretical terms, but we could always substitute statements about observables for ones about unobservables if any doubt were raised. For example, consider the theoretical concept of density. Every type of

material has a specific density, and we can explain why some bodies float in water and some do not by appeal to their densities. But the density of a thing is equal to its mass divided by its volume. If we can measure a thing's mass, on a scale, in a pan balance, or some other way, and we can measure its dimensions with a meter stick, we can calculate its density: That means we can "explicitly define" density in terms of mass and volume. In effect "density" is just an "abbreviation" for the quotient of mass and volume. Whatever we say about density we could say in terms of mass and volume. It may be more of a mouthful, but the empirical content of a claim about the mass of an object divided by its volume would be the same as the empirical content of any claim about its density. So, if we could explicitly define theoretical terms by way of observable ones, there would be no more trouble understanding what they mean than there is understanding what observable terms mean. There would be no chance of a theory introducing some pseudo-scientific term in a non-scientific theory that provides merely apparent explanatory power. Most important of all, we would know exactly under what observational conditions the things named by our observationally defined terms were present or not.

Unfortunately, hardly any of the terms that name unobservable properties, processes, things, states, or events are explicitly definable in terms of observable properties. Indeed, the explanatory power of theories hinges on the fact that their theoretical terms are *not* just abbreviations for observational ones. Otherwise, theoretical statements would simply abbreviate observational statements. And if they did that, then theoretical statements could summarize, but not explain, observational ones. Since density is by definition identical to mass divided by volume we could not appeal to their differing densities to explain why two objects of equal volume are of unequal mass; we would simply be repeating the fact that their ratios of mass to volume are unequal. More important, unlike "density," few theoretical terms can even be set equal to some finite set of observable traits or properties of things.

For example, temperature changes cannot be defined as equal to changes in the length of a column of mercury in an enclosed tube, because temperature also varies with changes in the length of a column of water in an enclosed tube, and changes in the resistance of an ohm-meter, or the shape of a bimetallic bar, or changes in the color of a heated object, etc. What is more, temperature changes occur even when there are no observable changes in the height of mercury or water in a tube. You cannot employ a conventional water or mercury thermometer to measure temperature changes smaller than about 0.1 degree centigrade, nor to measure temperatures that exceed the melting point of glass or fall below the freezing point of mercury or water or alcohol or whatever substance is employed. In fact there are some things whose temperatures change in ways that no thermometer we could currently design would record. So, some physical properties or changes in them do not seem to be observationally detectable. The situation for more theoretical properties than temperature is even murkier. If an "acid" is defined "as a proton donor" and no observations we can make give "empirical content" to the concept of a "proton donor" because we cannot

touch, taste, see, feel, hear, or smell a proton, then "acid" is a term with no meaning. On the other hand, we may define acid as "whatever turns blue litmus paper red," but then we won't be able to explain why some liquids do this and others don't.

Could we provide empirical meaning for the theoretical claims of science by linking complete theoretical statements with entire observable statements, instead of just individual theoretical terms with particular observable terms? Alas, no. The statement that the mean kinetic energy of the molecules in a particular gas container increases as pressure increases is not equivalent to any particular statement about what we can observe when we measure its temperature. This is owing to the fact that there are many different ways of measuring temperature observationally, and that using any one of them involves substantial further theoretical assumptions about the operation of thermometers, most especially the theoretical statement that absolute temperature at equilibrium equals mean kinetic energy.

The question we face cuts right to the heart of the problem about the nature of science. After all, the "official epistemology" of science is some form of empiricism, the epistemology according to which all knowledge is justified by experience: otherwise the central role of experiment, observation, and the collection of data in science would be hard to explain and justify. In the long run, scientific theorizing is controlled by experience: progress in science is ultimately a matter of new hypotheses that are more strongly confirmed than old ones as the results of empirical tests come in. Science does not accept as knowledge what cannot be somehow subject to the test of experience. But at the same time, the obligation of science to explain our experience requires that it go beyond and beneath that experience in the things, properties, processes, and events it appeals to in providing these explanations. How to reconcile the demands of empiricism with the demands of explanation is the hardest problem for the philosophy of science, indeed for philosophy as a whole. For if we cannot reconcile explanation and empiricism, it is pretty clear that it is empiricism that must be given up.

No one is going to give up science just because its methods are incompatible with a philosophical theory. We may have to give up empiricism for rationalism—the epistemology according to which at least some knowledge we have is justified without empirical test. But if some scientific knowledge is derived not from experiment and observation but, say, rational reflection alone, then who is to say that alternative world-views, myths, revealed religions, which claim to compete with science to explain reality, will not also claim to be justified in the same way?

The logical empiricist insists that we can reconcile empiricism and explanation by a more sophisticated understanding of how theoretical terms can have empirical content even though they are not abbreviations for terms that describe observations. Consider the concepts of positive and negative charge. Electrons have negative charge and protons positive ones. Now, suppose someone asks what the electron lacks that the proton has in virtue of which the former is said to have a negative charge and the latter is said to have a positive charge.

The answer of course is "nothing." The terms "positive" and "negative" used in this context don't represent the presence and absence of something. We could just as well have called the charge on the electron positive and the charge on the proton negative. These two terms function in the theory to help us describe differences between protons and electrons as they *manifest themselves in experiments* that we undertake with things we can observe. Electrons are attracted to the positive pole of a set of electrically charged plates and protons to the negative one. We can "see" the effects of this behavior in the visible tracks in cloud chambers or the gas bubbling up though the water in a chemical electrolysis set-up. The terms "positive" and "negative" make systematic contributions to the theory in which they figure, that are cashed in by the observational generalizations that the theory of atomic structure organizes and explains. The "empirical meaning" of the term "negative" is given by the systematic contribution that the term makes to the generalizations about what we can observe that follow from the assumptions of the theory about electrons being negatively charged. Remove the term from the theory, and the theory's power to imply many of these generalizations will be destroyed, the observations it can systematize and explain will be reduced. Whatever the extent of the reduction in explanatory power constitutes the empirical meaning of the term "negative."

We can use the same strategy to identify the empirical content of the term "electron" or "gene" or "charge" or any other term in our corpus of theories that names an unobservable thing or property in the same way. Each must make some contribution to the predictive and explanatory power of the theory in which it figures. To identify this contribution, simply delete the term from the theory and trace out the effects of the deletion on the theory's power. In effect, "charge" turns out to be defined "implicitly" as whatever it is that has the observable effects we lose when we delete the term "charge" from atomic theory, and similarly for any other theoretical term in any theory.

This in effect is the way in which theoretical terms are treated by the axiomatic approach to theories, the view that labeled itself "hypothetico-deductivism" and which we outlined in Chapter 7. Logical positivists sought to reconcile the explanatory power of the theoretical machinery of science with the constraints that observation places on science by requiring that legitimate theoretical terms be linked to observations through "**partial interpretation**." Interpretation is a matter of giving these terms empirical content, which may be quite different from the words scientists use to introduce them. Interpretation is partial because observations will not exhaust the empirical content of these terms, else they lose their explanatory power.

Another example may help. Consider the term "mass." Newton introduced this term with the definition "quantity of matter," but this definition is unhelpful because matter turns out to be as "theoretical" a notion as mass. Indeed, one is inclined to explain what matter is by appeal to the notion of mass, matter being anything that has any amount of mass. Mass is not explicitly defined in Newton's theory at all. It is an undefined term. Instead of being defined in the

theory, other concepts are defined by appeal to the concept of mass, for example, momentum, which is defined as the product of mass and velocity. But mass's empirical content is given by the laws in which it figures and their role in systematizing observations. Thus, mass is partially interpreted as that property of objects in virtue of which they make the arms of pan balances drop when placed upon them. We can predict that a mass coming into contact vertically with a pan balance will result in the balance arm moving because motion is the result of force, and force is the product of mass and acceleration, and moving a mass onto a pan balance causes the pan to have non-zero acceleration.

We should of course distinguish the "empirical meaning" of a term from its dictionary definition or semantic meaning. "Mass" is certainly a term with an English dictionary definition, even though its empirical meaning is quite different and in fact undefined in Newtonian mechanics.

So, the partial interpretation of mass is provided by the means we use to measure it. But these means do not define it. For one thing, the ways we measure mass by measuring its effects, like the motion of the arms of a pan balance, is something that mass causally explains. For another, there are many different ways of measuring mass by its effects, including some ways we may not yet have discovered. If such as-yet-undiscovered ways of measuring mass exist, then our interpretation of "mass" cannot be complete; it must be partial. And again, a complete interpretation in terms of observations would turn "mass" into an abbreviation for some set of observational terms, and would deprive it of its explanatory power.

The logical positivists advanced the claim that the unobservable terms of science need to be linked by meaning to observational terms, so that the really explanatory apparatus of science could be distinguished from pseudo-explanations that attempt to trade on the honorific title of science. Recall their interest in solving the demarcation problem discussed in Chapter 2. Ironically, the logical positivists were also the first to recognize that this requirement could not be expressed with the precision that their own standards of philosophical analysis required. Much of the history of positivism was devoted to framing what came to be known as a "principle of verification"—a litmus test that could be unambiguously applied to distinguish the legitimate theoretical terms of science from the illegitimate ones. Strong versions of the principle required complete translation of theoretical terms into observable ones. As we have seen, this requirement cannot be met by most of the terms invoked in scientific explanations; moreover we wouldn't want theoretical terms to satisfy this requirement because if they did so, they would lose their explanatory power with respect to observations.

The problem was that weaker versions of the principle of verification preserve the dross with the gold; they fail to exclude as meaningless terms that everyone recognizes as pseudo-scientific, and will not discriminate between real science and New Age psychobabble, astrology, or for that matter religious revelation. It is too easy to satisfy the requirement of partial interpretation. Take any pseudo-scientific term one likes and, provided one adds a general statement containing it to an already well-established theory, the term will pass

muster as meaningful. For example, consider the hypothesis that at equilibrium a gas is *bewitched* if its absolute temperature equals the mean kinetic energy of its molecules. Added to the kinetic theory of gases, this hypothesis makes the property of "being bewitched" into a partially interpreted theoretical term. And if one responds that the term "is bewitched" and the added "law" make no contribution to the theory, because they can be excised without reducing its predictive power, the reply will be made that the same can be said for plainly legitimate theoretical terms, especially when they are first introduced. What after all did the concept of "gene" add to our understanding of the distribution of observable hereditary characteristics in the decades before it was finally localized to the chromosome?

The demand that theoretical terms be linked to observations in ways that make a difference for predictions is far too strong a requirement; some theoretical terms, especially new ones, will not pass this test. It is also too weak a requirement, for it is easy to "cook up" a theory in which purely fictitious entities—vital forces, for example—play an indispensable role in the derivation of generalizations about what we can observe. If partial interpretation is too weak, we need to rethink the whole approach to what makes the unobservable terms of our theories meaningful and true or well justified, or even what makes coherent the claim that the unobservable things these terms name actually exist. What we need is a whole new theory of meaning to replace the empiricist theory inherited from the eighteenth century.

Scientific Realism vs. Antirealism

It may strike you that there is something about the way that logical empiricists treated this whole problem of the meaning of theoretical terms and the extent of our theoretical knowledge that gives it an artificial air. After all, though we may not be able to hear, taste, smell, touch, or see electrons, genes, quasars and neutron stars, or their properties, we have every reason to think that they exist. For our scientific theories tell us that they do, and these theories have great predictive and explanatory power. If the most well-confirmed theory of the nature of matter includes the laws about molecules, atoms, leptons, bosons, and quarks, then surely such things exist. If our most well-confirmed theories attribute charge, angular momentum, spin, or van der Waals forces to these things, then surely such properties exist. On this view theories must be interpreted literally, not as making claims whose meaning is connected to observations, but as telling us about things and their properties, where the meaning of the names for these things and their properties is no more or less problematical than the meaning of terms that name observable things and their properties, such as the Sun. And if this conclusion is incompatible with the theory of language enunciated above, which makes use of observational terms at the basement level of language and requires all other terms to be built out of them, then so much the worse for that theory of language. And so much the worse for a strict empiricist epistemology that goes along with it.

This approach to the problem of theoretical terms is widely known as "scientific realism," since it takes the theoretical commitments of science to be real, and not just (disguised) abbreviations for observational claims, or useful fictions that we create to organize these observations. (Notice that this use of the term "realism" contrasts with "Platonic realism," the claim that there are abstract objects—a very different sense than "scientific realism.")

Scientific realism couldn't get off the ground until the eclipse of positivism. For it reflects a question that logical positivists would have rejected as meaningless: Are the unobservable entities of physics real or not? Since they are unobservable, no observational test could decide this question and therefore any answer to the question was mere metaphysics. Give up the positivist criterion of verification though and the question becomes meaningful. And you would think the answer becomes obvious: of course there are atoms and protons, neutrons, electrons, photons, quarks, not to mention molecules, proteins, genes, etc. Yet surprisingly, this question is controversial and there is no obvious answer to it.

Whereas the logical positivists' starting point is a philosophical theory—empiricist epistemology—the scientific realist, or "realist" for short, starts with what realism takes to be a manifestly obvious fact about science: its great and ever-increasing predictive power. Over time our theories have improved in both the range and precision of their predictions. Not only can we predict the occurrence of more and more different kinds of phenomena, but over time we have been able to increase the precision of our predictions—the number of decimal places or significant digits to which our scientifically derived expectations match up with our actual meter readings. These long-term improvements translate into technological applications on which we increasingly rely, indeed on which we literally stake our lives every day. This so-called "instrumental success" of science cries out for explanation. Or at least the realist insists that it does. How can it be explained? What is the best explanation for the fact that science "works"? The answer seems evident to the realist: Science works so well because it is (approximately) true. It would be a miracle of cosmic proportions if science's predictive success and technological applications were just lucky guesses.

The structure of the scientific realist's argument is usually of the form:

1. P.
2. The best explanation of the fact that P, is that Q is true.
     Therefore,
3. Q is true.

Realists variously substitute for P the statement that science is predictively successful, or increasingly so, or that its technological applications are more and more powerful and reliable. For Q they substitute the statement that the unobservable things scientific theories postulate exist and have the properties that science attributes to them; or else the realist makes a somewhat weaker claim like "something like the unobservable entities that science postulates exist and have something like the properties that science attributes to them, and science is

ever-increasing its degree of approximation to the truth about these things and their properties." The structure of the argument from the truth of P to the truth of Q is that of an "**inference to the best explanation**." (These sorts of arguments are also known as "abductions" or "abductive arguments.")

This argument may strike the reader as uncontroversially convincing. It certainly appeals to many scientists. For they will themselves recognize the inference-to-the-best-explanation form of reasoning that the realist philosopher uses as one they employ in science. For example, how do we know that there are electrons and that they have negative charges? Because postulating them explains the results of the Millikan Oil Drop Experiment and the tracks in a Wilson Cloud Chamber.

But the fact that the argument form is used by scientists as well as philosophers to justify science is its Achilles heel. Suppose one challenges the argument for realism by demanding a justification for the inference form given in 1–3 above. The realist's argument aims to establish scientific theorizing as literally true or increasingly approximate to the truth. If the realist argues that the inference form is reliable because it has been used with success in science, the argument is potentially question-begging. The fact that an argument form has worked in the past cannot be used to underwrite confidence in its future, when the very question at issue is the idea that argument forms which have worked in the past will work in the future. In effect the realist argues that an inference to the best explanation's conclusion that scientific theorizing produces truths is warranted because science produces truths by using the inference form in question. To use an analogy we have used before, this is rather like backing up a promise to return a loan by promising to keep the promise to repay.

What is more, the history of science teaches us that many successful scientific theories have completely failed to substantiate the scientific realist's picture of why theories succeed. Well before Kepler, and certainly since his time, scientific theories have not only been false (and improvable), but if current science is any guide, they have sometimes been radically false in their claims about what exists and what the properties of things are, even as their predictive power has persistently increased.

One classical example is eighteenth-century phlogiston theory, which embodied significant predictive improvements over prior theories of combustion, but whose central explanatory entity, phlogiston, is nowadays cited with ridicule. Still another example is Fresnel's theory of light as a wave phenomenon. This theory managed substantially to increase our predictive (and explanatory) grasp on light and its properties. Yet the theory claims that light moves through a medium of propagation, an ether. The postulation of this ether is something one would expect in light of the difficulties traced above for the concept of gravity. Gravity is a mysterious force precisely because it doesn't seem to require any material through which it is transmitted. Without a medium of propagation, light waves would turn out to be as suspicious a phenomenon as gravity to the mechanistic materialism of nineteenth-century physics. Subsequent physics revealed that despite its great predictive improvements, the central theoretical

postulate of Fresnel's theory, the ether, does not exist. It is not required by more adequate accounts of the behavior of light. Postulating the ether contributed to the "unrealism" of Fresnel's theory. This at least must be the judgment of contemporary scientific theory. But by a "pessimistic induction" from the falsity—sometimes radical falsity—of predictively successful theories in the past, it would be safe to assume that our current "best-estimate" theories are also open to a similar fate. Since science is fallible, one might expect that such stories can be multiplied to show that over the long term, as science progresses in predictive power and technological application, the posits of its theories vary so greatly in their reality as to undermine any straightforward inference to scientific realism's interpretation of its claims.

What is more, scientific realism is silent on how to reconcile the knowledge it claims to have about the (approximate) truth of our theories about unobservable entities with the empiricist epistemology that makes observation indispensable for knowledge. In a sense, scientific realism is part of the problem of how scientific knowledge is possible, not part of the solution.

Some realists recognize the problematic character of our knowledge of the unobserved entities that theories refer to, combined with the track record of referential failure that underlies pessimistic induction. They propose to avoid both of these problems by adopting a view they call "structural realism." The succession of theories that characterize a predictively successful science do not constitute successively closer approximations to the truth about things in the world such as particles and fields. Rather, what they approach is the true mathematical structure of reality. These philosophers cite as a clear example of this sort of access to truths beyond observation the similarity of the mathematic formulations of Newtonian mechanics and its successors in quantum mechanics and general relativity. In fact Newton's laws can be derived deductively as special cases from quantum theory and from the theory of relativity. (Recall the Lorenz equation at the beginning of this chapter.) The reason is that these theories all share a core mathematical structure, for example, an inverse square equation. Structural realism declines to make any claim about the nature of the things that behave in accordance with these mathematical formulae that give common structure to theories that over time provide increasingly accurate predictions and explanations. They hold, however, that what scientific theories get right, even in the range of the unobserved, is the mathematical structure of reality, whence the name for their version of realism.

As an alternative to scientific realism, structural realism avoids some controversial commitments to knowledge of unobserved things. But it faces some evident problems: how to distinguish the mathematical structure or form of a theory from its content, and claims about particular things, expressed in statements that identify them via their (structural) properties and relationships, as theories commonly do. Without clear distinctions between mathematical form and factual content, it is hard to see how structural realism differs from scientific realism. Moreover, the question must arise whether knowledge of unobserved structure is easier to acquire or more easily reconciled with an empiricist epistemology than

is knowledge of unobservable things. To illustrate both of these objections: How can we tell that Newtonian mechanics, electrostatics, and the general theory of relativity are all correct that reality is approximated by an inverse square relation operating in their domains (the form of the theories), without having any understanding of what these domains are? What is more, structural realism needs to provide an independent criterion for what *the* mathematical form of a theory is. This can be a difficulty if theories have more than one form, or as yet no mathematical form, or are repudiated, superseded theories that nevertheless share some minimal form with approximately true ones. Every theory expressible in equations will share form at some level of abstraction. Which is the level at which structural realism argues that the preservation of form across successive theories reveals the nature of unobserved reality?

There is an alternative to scientific realism, that is much more sympathetic to empiricism, which has long attracted some philosophers and scientists. It bears the title "**instrumentalism**." This label names the view that scientific theories are useful instruments, heuristic devices, tools we employ for organizing our experience, but not literal claims about the world that are either true or false. This philosophy of science goes back at least to the eighteenth-century British empiricist philosopher George Berkeley, and is also attributed to leading figures of the Inquisition who sought to reconcile Galileo's heretical claims about the motion of the Earth round the Sun with holy writ and papal pronouncements. According to some versions of the story, these learned churchmen recognized that Galileo's heliocentric hypothesis was at least as powerful in prediction as Ptolemaic theories, according to which the Sun and planets moved around the Earth; they accepted that it might be simpler to use in calculations of the apparent positions of the planets in the night sky. But the alleged motion of the Earth was observationally undetectable—it does not feel to us that the Earth is moving. Galileo's theory required that we disregard the evidence of observation, or heavily reinterpret it. Therefore, these officers of the Inquisition urged Galileo to advocate his improved theory not as literally true, but as more useful, convenient, and effective an instrument for astronomical expectations than the traditional theory. Were Galileo so to treat his theory, and remain silent on whether he believed it was true, he was promised that he would escape the wrath of the Papal Inquisition. Although at first he recanted, Galileo eventually declined to adopt an instrumentalist view of the heliocentric hypothesis and spent the rest of his life under house arrest.

Subsequent instrumentalist philosophers and historians of science have suggested that the Church's view was more reasonable than Galileo's, albeit for all theories and not just this one. And although Berkeley did not take sides in this matter, his arguments from the nature of language (sketched above) to the unintelligibility of realism (and of realistic interpretations of parts of Newton's theories) made instrumentalism more attractive. Berkeley went on to insist that the function of scientific theorizing was not to explain but simply to organize our experiences in convenient packages. On this view, theoretical terms are not abbreviations for observational ones, but more like mnemonic devices, acronyms,

uninterpreted symbols without empirical or literal meaning. And the aim of science is constantly to improve the reliability of these instruments, without worrying about whether reality corresponds to them when interpreted literally.

It is worth noting that the history of the physical sciences from Newton onward shows a cyclical pattern of succession between realism and instrumentalism among scientists themselves. The realism of the seventeenth century, the period in which mechanism, corpuscularianism, and atomism held sway, was succeeded in the eighteenth century by the ascendancy of instrumentalist approaches to science, motivated in part by the convenient way that instrumentalism dealt with Newton's mysterious force of gravity. By treating his theory of gravity as merely a useful instrument for calculating the motion of bodies, it could ignore the question of what gravity really is. By the nineteenth century, with advances in atomic chemistry, electricity and magnetism, the postulation of unobservable entities returned to favor among scientists. But then it again became unfashionable in the early twentieth century as problems arose for the realist's interpretation of quantum mechanics as a literally true description of the world began to mount. On the standard understanding of quantum mechanics, electrons and photons seem to have incompatible properties—being both wave-like and particle-like at the same time, and neither seem to have physical location until observed by us. And providing an interpretation of quantum mechanics that solves these interpretational problems would make no contribution to its already phenomenally great predictive power. These are reasons why it is more than tempting to treat quantum mechanics as a useful instrument for organizing our experience in the atomic physics lab, but not as a set of true claims about the world independent of our observation of it.

How does instrumentalism respond to the realists' claim that only realism can explain the success of science? The instrumentalist replies quite consistently with the following argument: that any explanation of the success of science that appeals to the truth of its theoretical claims either advances our predictive powers with respect to experience or it does not. If it does not, then we may neglect it and the question it purports to answer as without empirical significance. If, on the other hand, such an explanation would enhance the usefulness of our scientific instruments in systematizing and predicting experience, then instrumentalism can accept the explanation as confirming its treatment of theories as useful instruments, instead of descriptions of nature.

Some philosophers have sought a compromise between instrumentalism and realism that will allow us to take our theories at face value while avoiding commitments that empiricism renders problematic. These compromises are attempts to have one's cake and eat it too: we agree with the scientist that scientific theories do purport to make claims about the world and especially about the unobservable underlying mechanisms that explain observations, and we can agree with the instrumentalist that knowledge of such claims is impossible. But we may argue that the objective of science should be nothing more or less than systematizing experience. Therefore we can be agnostic about whether scientific theories are true, approximately true, false, convenient fictions or whatever, just

so long as they enable us to control and predict the phenomena. We can and should accept them, without of course believing them (that would be to take a position on their truth). Science should be content with simply predicting, with increasing precision, an ever-wider range of our experiences. In short, scientists should aim at what the instrumentalist recommends without embracing the instrumentalists' reason for doing so. Science certainly is an instrument. It's just that we cannot tell whether it is more than an instrument. And for all purposes it is enough that scientific theory be "empirically adequate." Recalling the words of the seventeenth-century natural philosophers, on this view, all we should demand of science is that it should "save the phenomena."

This combination of a realist interpretation for the claims of theoretical science with an instrumentalist epistemology has been called "**constructive empiricism**" by its developer, Bas van Fraassen. Few philosophers and fewer scientists will consider constructive empiricism to be an enduring stable equilibrium in the philosophy of science. After all, if science is either (increasingly approximately) true or (persistently) false in its representation of the world, but we can never tell which, then the treatment of science as a description of reality just drops out of intellectual matters. If we cannot tell which of these exhaustive and exclusive alternatives applies, then whichever does is probably irrelevant, as the instrumentalist claims. On the other hand, if we must forever withhold our judgment about the truth of the most predictively powerful and technologically successful body of hypotheses we can formulate, then the epistemological question of whether we can have scientific knowledge becomes as irrelevant to science as the skeptic's question of whether I am now dreaming or not.

Realism, instrumentalism, and constructive empiricism approach the problem of theoretical entities and the terms that name them with the same two assumptions in common: They are predicated on the assumption that we can divide up the terms in which scientific laws and theories are expressed into observational and non-observational (or theoretical) ones; all three agree that it is our knowledge of the behavior of observable things and their properties that tests, confirms, and disconfirms our theories. For all three, the court of last epistemological resort is observation. And yet, as we shall see below, exactly how any one observation tests *any* part of science, theoretical or not, is no easy thing to understand.

The logical positivists had no patience with the realist–instrumentalist dispute about scientific theories, because it was not one that could be subjected to empirical test. Other philosophers have condemned it as reflecting profound misunderstandings about science, about truth, and about knowledge. According to Arthur Fine, the entire dispute may be avoided, for example, by adopting the *Natural Ontological Attitude* towards scientific theories. In ordinary life we treat many propositions as true without inquiring closely into what their truth commits us to—their ontological commitments. We naturally assent to such statements as that "2 is the only even prime." But if someone demanded of us whether we believe that there is a thing that is named by the numeral "2," and whose existence is required for the truth of the statement, many people would

demur. If asked what a number is, if not a particular thing with concrete existence, few people would give an answer. Those who reply that the number 2 is an idea in the mind, would soon be convinced of the unsatisfactory character of this hypothesis. But none of this has any tendency to weaken our commitment to the truth of the statement that "2 is the only even prime." All this is quite natural. So too in science, we can, we do, and we should accept the truth of those theories we adopt without taking sides on matters in dispute between proponents of scientific realism and **antirealism**, without accepting or rejecting the warrant of inferences to the best explanation generally, or the no-miracles argument, or pessimistic induction. In particular the Natural Ontological Attitude can remain agnostic about the long-term progress of science, about whether Newtonian mechanics and the special theory of relativity mean the same thing by mass, and whether the corpuscularianism of the seventeenth century is vindicated by the atomic theory of the twentieth century. In ordinary life we accept that the best available explanation may not be true, but is still explanatory; even if realism is the best explanation for the predictive improvements of science, we need not conclude that it is true, or for that matter that successive theories in science are closer and closer approximations to the truth.

The realist and the antirealist will be pardoned for wondering whether the Natural Ontological Attitude misses the point. As a report about how scientists discuss their theories and take sides on them, Fine's approach seems unexceptional. But philosophers will rightly reject a description of what scientists do that is absolutely neutral between logically incompatible claims. We can't be neutral as between treating its claims as purporting to be truths about the world versus treating them as useful instruments that it is convenient to characterize as descriptions. Once we fix the meaning of terms like truth and reference, the questions that realists and antirealists dispute about science are real and unavoidable. In fact, the same must be said about "2 is the only even prime." Mathematicians may not be disturbed about what numbers are, having left the question to philosophy. But the question remains to be answered. The same goes for how we know about the existence of unobservable concrete objects.

Summary

The axiomatic account of scientific theories explains how the laws of a theory work together to provide an explanation of a large number of empirical or observable regularities by treating theories as deductively organized systems, in which the assumptions are hypotheses confirmed by observations about the generalizations that are derived from them. This conception of laws as hypotheses tested by the consequences deduced from them is known as "hypothetico-deductivism," a well-established account of how theories and experience are brought together.

Theories often explain by identifying the underlying unobserved processes or mechanisms that bring about the observable phenomena that test the theories. Reductionism labels a longstanding view about the relationship of scientific

theories to one another. According to reductionism, as a science deepens its understanding of the world, narrower, less accurate, and more special theories are revealed to be special cases of or explainable by derivation from broader, more complete, and more accurate theories. Derivation requires the logical deduction of the axioms of the narrower theory from the broader one, and often the correction of the narrower theory before the deduction is effected. Reductionists seek to explain the progress of science over the period since the Newtonian revolution by appeal to these intertheoretical relations. The reduction of scientific theories over centuries, which seems to preserve their successes while explaining their failures (through correction), is easy to understand from an axiomatic perspective on scientific theories.

The hypothetico-deductivism of the axiomatic account of theories (and indeed the general epistemological perspective of science as based on observation and experiment), however, faces grave difficulty when it attempts to explain the indispensability of terms in theories that identify theoretical, unobservable entities, like cellular nuclei, genes, molecules, atoms, and quarks. For, on the one hand, there is no direct evidence for the existence of the theoretical entities these terms name, and, on the other hand, theory cannot discharge its explanatory function without them. Some theoretical entities, such as gravitation, are truly troublesome, yet at the same time we need to exclude from science all mysterious and occult forces for which no empirical evidence can be provided. The notion that meaningful words must eventually have their meaning given to them by experience is an attractive one. Yet finding a way for theoretical language to pass this test, while excluding the terms of uncontrolled speculation as meaningless, is a challenge that any account of scientific theories must face.

The puzzle that hypothesizing theoretical entities is indispensable to explanation and unregulated by experience is sometimes solved by denying that scientific theories seek to describe the underlying realities that systematize and explain observational generalizations. This view, known as instrumentalism, treats theory as a heuristic device, a calculating instrument for predictions alone. By contrast, realism, the view that we should treat scientific theory as a set of literally true or false descriptions of unobservable phenomena, insists that only the conclusion that a theory is approximately true can explain its long-term predictive success. Instrumentalists controvert this explanation.

Study Questions

1. Defend or criticize: "Where theories cannot be expressed mathematically reduction is not possible."
2. Why would someone hold this view: "Real knowledge requires truth. This means that in science we have to choose between our scientific explanation and knowledge"?

3. Is "constructive empiricism" really a viable middle course between instrumentalism and realism?

4. Evaluate the following argument for realism: "As technology progresses, yesterday's theoretical entities become today's observable ones. Nowadays we can detect cells, genes, and molecules. In the future we will be able to observe photons, quarks, etc. This will vindicate realism."

5. Does instrumentalism owe us an explanation of the success of science? If so, what is it? If not, why not?

6. Is "structural realism" a version of realism at all? Does it accord concrete reality to abstract structures?

Suggested Readings

A statement of the post-positivist view of reduction by Nagel is to be found in Curd and Cover, along with an introduction by Nickles, "Two Concepts of Intertheoretical Reduction," and Kitcher's essay on reduction in biology, "1953 and All That: A Tale of Two Sciences."

The view of scientific progress reflected in the positivist notion of reduction is examined in W. Newton-Smith, *The Rationality of Science*, M. Spector, *Concepts of Reduction in Physical Science*, and A. Rosenberg, *The Structure of Biological Science*. But many papers have been written and continue to appear on this issue, especially in the journals *Philosophy of Science* and the *British Journal for Philosophy of Science*. P. Feyerabend's vigorous attack on the complacent picture of progress as reduction, "Explanation, Reduction, and Empiricism," reprinted in Balashov and Rosenberg, has been very influential, especially when harnessed together with some interpretations of Thomas Kuhn's views, as we shall see in Chapter 12. Another of Feyerabend's papers on reduction appears in the Curd and Cover anthology. Kitcher, "Theories, Theorists, and Theoretical Change," offers a sophisticated discussion of theoretical continuity through replacement, with particular reference to the case of phlogiston and oxygen. This paper too is reprinted in Balashov and Rosenberg, and treats matters also taken up in Chapter 6.

Hempel's paper "The Theoretician's Dilemma," in *Aspects of Scientific Explanation*, expresses the problem of reconciling the indispensability of theoretical entities for explanation with the empiricist demand that the terms naming those entities be observationally meaningful. Other papers in *Aspects*, including "Empiricist Criteria of Significance: Problems and Changes," reflect these problems. These papers are both reprinted in Lange's anthology.

Among the earliest and most vigorous post-positivist arguments for realism is J. J. C. Smart, *Between Science and Philosophy*. Van Fraassen's *The Scientific Image* argues for constructive empiricism. The debate between realists, antirealists, and instrumentalists is well treated in J. Leplin (ed.), *Scientific Realism*, which

includes papers defending realism by R. Boyd and E. McMullin, a development of the "pessimistic induction" from the history of science to the denial of realism by L. Laudan, a statement of van Fraassen's "constructive empiricism," and a plague on both realism and antirealism pronounced by Arthur Fine, "The Natural Ontological Attitude." Van Fraassen's views are more fully worked out in *The Scientific Image*. J. Leplin, *A Novel Argument for Scientific Realism*, is a more recent defense of realism against van Fraassen and others. P. Churchland and C. A. Hooker (eds.), *Images of Science: Essays on Realism and Empiricism*, is a collection of essays discussing "constructive empiricism." Laudan's arguments against realism are powerfully developed in "A Confutation of Convergent Realism," reprinted in Balashov and Rosenberg, which also includes an illuminating discussion of van Fraassen's views and realism by Gutting, "Scientific Realism vs. Constructive Empiricism: A Dialogue," and a historically informed defense of realism, E. McMullin, "A Case for Scientific Realism." Curd and Cover also reprint several important papers on the realism/antirealism controversy, including one of the first, by Grover Maxwell, along with papers by van Fraassen, Laudan, and Fine. Lange's anthology reprints John Worral's important paper, "Structural Realism: The Best of Both Worlds," along with an exposition by van Fraassen of constructive empiricism. Ladyman and Ross offer a sustained argument for a radical ontological structural realism in *Everything Must Go*. P. Kyle Stanford, *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*, advances new arguments against realism.

Nancy Cartwright and Keith Ward, *Rethinking Order: After the Laws of Nature*, articulates a radical alternative to both realism and instrumentalism about theory and its relation to the world.

The problem of the underdetermination of theory by evidence has encouraged a number of philosophers to invoke the importance of multiple incompatible theories and explanations driven by them as an enduring feature of science. These philosophers of science adopt the label *pluralism* for their view. Pluralists argue that underdetermination reveals that disciplines in which only one explanatory paradigm or theoretical tradition obtains are often hindered by the social and political exclusion of alternative approaches. They argue that disciplines in which larger numbers of competing models and theories are encouraged, and critically assessed, advance the understanding of their domains more effectively, even when underdetermination prevents a significant narrowing of alternatives. Pluralists have made an important contribution to the debate about the objectivity of science even in the shadow of underdetermination. They have shown that a discipline conscious of the social forces that shape research agendas and the assumptions that drive it is more likely to remain objective by identifying these factors that drive incompatible models, theories, and explanations. The most forceful articulation of this sort of scientific pluralism is to be found in the work of Helen Longino, *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*, and *The Fate of Knowledge*.

9   Theory Construction vs. Model Building

Overview

In many disciplines, scientists increasingly describe the product of their research not as theories but as **models**. In some disciplines it is clear that building a sequence of models is a process that is expected to culminate in a grand or at least a more general theory. In other disciplines, the aim of research is a model, and a theory is in fact a set of models. Moreover, while theories in science are thought of as general scientific hypotheses that the scientist considers good candidates for being laws of nature with explanatory and predictive import, none of these things can in general be said of models. They are not necessarily offered as scientists' best guesses about laws, nor even necessarily with the intention of explaining or predicting any actual experimental or other observable processes.

All this suggests that the logical positivist and post-positivist absorption with theories as axiomatic systems composed of laws and regularities may not fit either as a description or a rational reconstruction of the theoretical activities of scientists. It may also mean that some or all of the philosophical problems arising out of the nature of scientific theories may be circumvented by an approach that focuses on models as the units of scientific research. In this chapter we explore some of these issues.

Biology is one discipline in which there seems to be at most one theory—Darwin's theory of natural selection—and lots of models about phenomena at every level of biological organization from enzymes to populations. This makes it convenient to explore the relationship between theories and models as it plays out in biology with some detailed examples.

Theories and Models

The post-positivist view of theories, hypothetico-deductivism, treated theories in all the sciences as sets of assumed or unproved axioms from which theorems are derived. The axioms, the underived, basic laws in the theory, are usually expressed in terms that make no reference to observations and so cannot be tested directly. However, the logically derived theorems are couched in observational terms that enable scientists to test them directly, and which can be systematically linked to the axioms and through them to the theoretical terms in the axioms, thus giving them meaning. In Chapter 8 we noted several difficulties with this axiomatic approach. One was the difficulty of deriving narrower theories from broader, more basic ones, and earlier theories from later, more fundamental ones. Another was the difficulty of providing a satisfactory account of the meaning (fullness) of theoretical terms. Once the empiricist idea that their meaning was given by indirect connection—partial interpretation—to observational terms, the problem of scientific realism becomes a significant one.

Nevertheless, axiomatization is a powerful notion that does seem to provide an account of the systematic explanatory power of theories. Axiomatization is plainly not the way in which scientists actually present their theories. Most of the philosophers who advocated it as giving the structure of theories didn't suppose any such thing. They treated it as a rational reconstruction of the ideal or essential nature of a scientific theory, which explains how it fulfills its function. But there are two immediate and related problems that the axiomatic model faces.

The first is that there is no role in the axiomatic account for models as they actually figure in the sciences. Yet nothing is more characteristic of theoretical science than its reliance on the role of models. Positivist and post-positivist philosophers of science used the word "model" in the way mathematicians do, meaning an interpretation of an abstract axiomatic system. This has almost nothing to do with the meaning of the word "model" in science. Consider the planetary model of the atom, the billiard-ball model of a gas, Mendelian models of genetic inheritance, the Keynesian macro-economic model. Indeed, the very term "model" has supplanted the word "theory" in many contexts of scientific inquiry. This is especially the case in the so-called "special sciences," where models have supplanted *ceteris paribus* laws. It is pretty clear that often the use of this term suggests the sort of tentativeness that the expression "just a theory" conveys in non-scientific contexts. But in some domains of science there seem to be nothing but models, and either the models constitute the theory or there is no

separate thing at all that is properly called a theory. This is a feature of science that the axiomatic approach must explain or explain away.

The second problem for the axiomatic approach is the very idea that a theory is an axiomatized set of sentences in a formalized mathematical language. The claim that a theory is an axiomatic system is in immediate trouble in part because, as we noted above, there are many different ways to axiomatize the same set of statements. But more than that, a particular axiomatization is essentially a linguistic thing: it is stated in a particular language, with a particular vocabulary of defined and undefined terms, and a particular syntax or grammar. Now ask yourself, is Euclidean geometry correctly axiomatized in Greek, with its non-Roman alphabet, or nineteenth-century German with its gothic letters, its verbs at the end of sentences and its nouns inflected, or in English, or in Mandarin pictograms? The answer is that Euclidean geometry is indifferently axiomatized in any language in part because it is not a set of sentences in a language but a set of propositions that can be expressed in an indefinite number of different axiomatizations in an equally large number of different languages. To confuse a theory with its axiomatization in a language is like confusing the number 2—an abstract object—with the concrete inscriptions, like "dos," "II," "zwei," "$10_{(\text{base 2})}$," that we employ to name it. Confusing a theory with its axiomatization is like mistaking a proposition (again, an abstract object) for the particular sentence (a concrete object) in a language used to express it. "Es regnet" is no more the proposition that it is raining than "Il pleut," nor is "It's raining" the correct way to express it either. All three of these inscriptions express the same proposition about the weather, which is not itself in any language. Similarly, we may not want to identify a theory with its axiomatization in any particular language, not even in some perfect, mathematically powerful, logically clear language. And if we don't want to do this, the axiomatic account is in some difficulty, to say the least.

What is the alternative? Let's start with models for the phenomena that scientists actually develop, for example the Mendelian model of hereditary transmission. Biologists simply treat the two regularities that Mendel discovered in the 1860s (the "laws" of independent assortment of genes for heredity traits and of the segregation of genes at meiosis) as components of a Mendelian model. Both of these laws are known to have exceptions, and like other laws in the "special sciences" need to be treated as embodying implicit or explicit *ceteris paribus* clauses. But they are sufficiently well confirmed and practically useful that biologists are prepared to treat them as defining a Mendelian model.

This practice is common in social science as well. In 1936 John Maynard Keynes published *The General Theory of Employment, Interest and Money*. This long book was subject to controversial interpretations and embodied few equations in a discipline—economics—that was becoming more quantitative. Within ten years many economists had agreed on a set of three linear equations, the so-called Keynesian model, to express the core content of the theory. An economy is Keynesian to the extent that it satisfied these equations, which were treated as individually necessary and jointly sufficient for an economy being Keynesian.

Y = C + I + g   [aggregate income = consumption + investment + government
                       spending]
C = f(Y)   [consumption is a function of income]
I = E(R)   [investment is a function of "the marginal efficiency of capital"]

In this model, the first equation is intended to be true by definition; the second to be obvious and of interest only when the value of the function f can be given; while the third will obtain only in a perfectly competitive capital market. So, none of the components of the Keynesian model is really to be understood as a robust contingent truth of the sort that figures as the axiom of some logical system that purports to be a scientific theory.

Similarly, we may express the model for a Newtonian system: A Newtonian system is any set of bodies that behave in accordance with the following two formulae: $F = Gm_1m_2/d^2$—the inverse square law of gravitational attraction, $F = ma$—the law of free-falling bodies, and the laws of rectilinear motion and the law that for every action there is an equal and opposite reaction (or the more fundamental law of the conservation of momentum). Again, these four features define a Newtonian system. Now, let's consider what arrangement of things in the world satisfies these definitions? Well, by assuming that the planets and the Sun are a Newtonian system, we can calculate the positions of all the planets with great accuracy as far into the future and as far into the past as we like. So, the solar system satisfies the definition of a Newtonian system. Similarly, we can calculate eclipses—both solar and lunar—by making the same assumption for the Sun, the Earth and the Moon. And of course we can do this for many more sets of things—cannon balls and the Earth, inclined planes and balls, pendula, and for that matter comets, double stars and galaxies. In fact if we assume that gas molecules satisfy our definition of a Newtonian system, then we can predict their properties too.

The definition given above for a Newtonian system is not the only definition we could give. And it might even be preferable to adopt another definition, if for example the alternative definition could avoid some of the problems that bedevil the textbook version of Newton's theory, especially its commitment, in the inverse square law, to a force that can be transmitted at infinite speed through a perfect vacuum and from which nothing can be shielded, i.e. gravity. The highly creative Nobel Prize-winning physicist Richard Feynman advanced an alternative formulation for Newton's theory which substitutes for the inverse square law a formula that gives the gravitational force at a point in space as the function of the average of the gravitational forces on other points surrounding that point: Phi = average Phi − Gm/2a, where Phi is the gravitational potential or force at any given point, a is the radius of the surrounding sphere on the surface of which the average gravitational force, average Phi, is calculated, G is the same constant as figures in the formula above, and m is the mass of the objects at the point on which gravity is exerted. Feynman in fact noted that one may prefer this formula to the usual one because $F = Gm_1m_2/d^2$ suggests that gravitational force

operates over large distances instantaneously, whereas the less familiar equation gives the values of gravitational force at a point in terms of values at other points which can be as close as one arbitrarily chooses. But either definition will work to characterize a Newtonian gravitational system.

Now the reason we call these definitions models is that they "fit" some natural processes more accurately than others; they are often deliberate simplifications which neglect causal variables that we know to exist but are small compared to the ones the model mentions. Moreover, even when we know that things in the world don't really fit them at all, they may still be useful calculating devices, or pedagogically useful ways of introducing a subject. Thus, a Newtonian model of the solar system is a deliberate simplification that ignores friction, small bodies like comets, moons and asteroids, and electric fields, among other things. Indeed we know that the model's exact applicability is disconfirmed by astronomical data on, for example, Mercury's orbit. And we know that the model's causal variable does not really exist (there is no such thing as Newtonian gravity which acts at a distance; rather space is curved). Nevertheless, it is still a good model for introducing mechanics to the student of physics and for sending satellites to the nearest planets. (Despite the fact that Einstein's model had replaced Newton's nearly 60 years earlier, the Apollo 11 Moon landing mission in 1969 was performed entirely using Newtonian calculations.) Moreover, the advance of mechanics from Galileo and Kepler to Newton and Einstein is a matter of the succession of models, each of which is applicable to a wider range of phenomena and/or more accurate in its predictions of the behavior of the phenomena.

Often a model advanced by a scientist is true by definition. When it isn't, it is usually because the scientist is not interested in its truth or falsity in general. Rather, what is interesting is which systems it is true about, and under what general conditions it does not apply. Thus, for example, an ideal gas is by definition just what behaves in accordance with the ideal gas law. The empirical or factual question about a model is whether it "applies" to anything closely enough to be scientifically useful—to explain and predict its behavior. Thus, it will be a hypothesis that the Newtonian model applies well enough to, or is sufficiently well satisfied by the solar system. Once we specify what "well enough" or "sufficiently well satisfied" entails, this is a hypothesis that usually turns out to be true. The unqualified claim that the solar system is a Newtonian system is, we know, strictly speaking false. But it is much closer to the truth than any other hypothesis about the solar system except the hypothesis that the solar system satisfies the model propounded by Einstein in the general theory of relativity. And a theory? A theory is a set of hypotheses claiming that particular sets of things in the world are satisfied to varying degrees by a set of models that reflect some similarity or unity. This will usually be a set of successively more complex models. For example, the kinetic theory of gases is a set of models that begins with the ideal gas law we have seen before, $PV = nRT$. This model treats molecules as billiard balls without intermolecular forces and assumes they are mathematical points. The theory includes a subsequent improvement due to van der Waals, $(P + a/V^2)(V − b) = rT$, in which "a" represents the intermolecular

forces and "b" reflects the volume molecules take up, both neglected by the ideal gas law. And there are other models, Clausius' model and ones that introduce quantum considerations, as well.

Semantic vs. Syntactic Approaches to Theories and Models

Proponents of this approach to theories, according to which they are sets of models (that is of formal definitions, along with claims about what things in the world satisfy these definitions), call their analysis the "semantic" account of scientific theories and contrast it to the axiomatic account, which they call the "syntactic" account. There are two related reasons: (1) it requires derivation of empirical generalizations from axioms in accordance with rules of logic, which are the syntax of the language in which the theory is stated; (2) the derivations that logical rules permit operate on the purely formal features—the syntax—of the axioms, and not the meaning of their terms. Notice that although models will be identified by linguistic items on the **semantic** view, definitions, hypotheses, and theories will not be linguistic items. They will be (abstract) propositions expressible in any language, to the effect that the world or some part of it satisfies to some degree or other one or more models, expressed indifferently in any language convenient for doing so.

But surely this is not the chief advantage of the semantic view. For after all, the axiomatic account may best be understood as the claim that a theory is a set of axiom systems in any language that express all the same propositions as axioms or theorems, or that it is the set of all such axiom systems that best balance simplicity and economy of expression with power in reporting these propositions. If the linguistic or non-linguistic character of theories is a problem, it is a rather technical one for philosophers, which should have little impact on our understanding of scientific theories. The advantage of the semantic over the syntactical approach to theories must lie elsewhere.

One advantage that the semantic approach has is that it focuses attention on the role and importance of models in science in a way that the axiomatic account does not. In particular, it is hard for axiomatic analysis to accommodate the formulation of models known from the outset to be false but useful idealizations. It won't do simply to interpret PV = nRT not as a definition of an ideal gas, but as an empirical generalization about real objects to be derived from axioms of the kinetic theory of gases, if we know that the statement could not be true. We don't want to be able to derive such falsehoods directly from our axiomatic system, for such derivations imply that one or more of the axioms is false. What we may want is to find a place for models within an axiomatic approach.

A related advantage of the semantic approach is often claimed. In some areas of science, it is sometimes felt that there is no axiomatization available for the relevant laws, or that axiomatization would be premature and freeze the development of ideas that are still being formulated. To suggest that thinking in a discipline can or should be rationally reconstructed as an axiomatization would therefore be disadvantageous. Sometimes it is claimed that evolutionary theory

in biology is like this—still too fluid a subject to be formalized into one canonical expression of its contents. When we try to frame the theory of natural selection into an axiomatic system, the result is often rejected by evolutionary biologists as failing to adequately reflect the full richness of Darwin's theory and its latter-day extensions. We explore these matters in detail in the next section.

Meanwhile, can particular sciences or sub-disciplines really remain agnostic about whether there are fundamental underlying theories towards which models in their disciplines are moving? They must do so, if there simply is no set of higher-level general laws in the discipline that explains lower-level regularities, and their exceptions. Recall one of the metaphysical attractions of the axiomatic approach: its commitment to axiomatization as an account of how a theory explains by uncovering underlying mechanisms. Consider the metaphysical thesis that at bottom the universe is simple in composition and operation, and all the diversity of more complicated and more composite things is the result of simplicity at the bottom. This thesis suggests that there is a true theory about layers of causal laws, each of which rests on a more fundamental layer of a smaller numbers of laws about a smaller range of simpler objects that imply the less fundamental laws. It is a short step to the conclusion that there should be a uniquely correct axiomatization of this theory that reflects the structure of reality. The logical empiricists who first advanced the axiomatic account would not have expressed such a view, because of their desire to avoid controversial metaphysical debate. Philosophers less averse to metaphysics will certainly find the view a motivation for adopting a syntactic model of theories. By contrast, philosophers who reject this metaphysical picture have a concomitant reason to adopt the semantic approach to theories. This approach makes no commitments to any underlying simplicity or to the reducibility of less fundamental theories (i.e. sets of models) to more fundamental theories (i.e. sets of more fundamental models). If nature is just not simple, the structure of science will reflect this fact in a plethora of sets of models, and a dearth of axiomatic systems. And it will encourage instrumentalism about the character of theories and their claims about reality.

Notice that the instrumentalist can refuse to be a party to this debate about whether theories describe reality. For the instrumentalist must be indifferent to the question of whether there is some set of laws that explain why the models work. Indeed, so far as instrumentalism is concerned, models might just as well supplant theory altogether in the advancement of science. Who needs theory if it cannot provide greater empirical adequacy than the models whose success it explains? It is for this reason that it is sometimes supposed that the semantic view of theories is more amenable to an instrumentalist philosophy of science than the syntactic or axiomatic approach. Not only will the instrumentalist be sympathetic to the semantic approach's indifference to the controversy surrounding realism, but taking the construction of models as the fundamental task of science may also enable the philosopher of science to avoid questions about the nature and existence of laws, the "real" character of causation, and the entirely metaphysical issue of what causal, nomic, natural, or physical necessity consists in.

By contrast realists cannot and do not wish to avoid these issues. They hold that both the success and especially the increasing accuracy of the succession of models in these sub-disciplines demands explanation. Of course, some may argue that it is possible for a set of models in, say, evolutionary biology, to provide considerable predictive power and indeed increasing precision, even though the only general theory in biology is to be found at the level of molecular biology. For example, it might turn out that the biological models we formulate work for creatures with our peculiar cognitive and computational limitations and practical interests, but that the models don't really reflect the operation of real laws at the level of organization of organisms and populations of them. This would be a realist's explanation for the absence of laws at some levels of organization where there are effective models. But the realist cannot adopt such a stratagem to explain away the absence of laws that might explain the success of models in physics or chemistry.

Moreover, the realist will argue, the semantic approach shares with the axiomatic account a commitment to the existence of theories that are distinct from and different from the models on which it focuses. For the semantic approach tells us that a theory is the substantive claim that a set of models which share some features in common are satisfied by things in the world. A theory is the set of definitions that constitute the models *plus* the claim that there are things that realize, satisfy, instantiate, and exemplify these definitions sufficiently well to enable us to predict their behavior (observable or unobservable) to some degree of accuracy. Applying a model to real processes is an *ipso facto* commitment to the truth of this substantive claim. But such a claim is more than a mere instrument or useful tool that enables us to organize our experiences. Accordingly, like the axiomatic account, the semantic approach is committed to the truth of general claims in science. And the semantic view of theories has all the same intellectual obligations to explain why theories are true or approximately true, or at least moving successively closer to the truth that the axiomatic account does.

Additionally, the semantic view of theories may face the same problems as those with which we left the axiomatic account at the end of the last chapter. Many of the models in science are definitions of unobserved, theoretical systems, such as the Bohr model of the atom. Therefore, the semantic view of theories faces the same problem of reconciling empiricism with the indispensability of theoretical terms, or equivalently the commitment to theoretical objects, that the axiomatic account does. Applying a model to the world requires that we connect it to what can be observed or experienced, even if what is observed is a photograph that we interpret as representing a sub-atomic collision, or a binary star, or the semi-conservative replication of a DNA molecule. Whether the theory (or a model) explains data as the realist holds, or only organizes it as the instrumentalist holds, the theory can do neither without recourse to claims about this realm of unobservable things, events, processes, or properties that an empiricist epistemology makes problematic. But the final epistemic arbiter for science is observation. And yet, as we shall see below, how observation tests any part of science, theoretical or not, is no easy thing to understand.

Quite apart from this dispute about whether a theory-free science that deals only in models can do justice to the descriptive or representational tasks of science, proponents of the centrality of models rightly note that there is more to the aims of science and scientists than giving a uniquely correct, or even one among several correct, descriptions of the world. Often what scientists want or need is not a linguistic description but a physical model—like Watson and Crick's tinkertoy model of DNA, which was far more suggestive for understanding genetics than its verbal description. Sometimes physicists prefer a mathematical model that cannot be interpreted or expressed coherently in any language. This seems to be an attitude widespread in quantum mechanics. Their interests are purely predictive and their models are never assessed as representations of reality. In other cases, scientists construct models of completely imaginary circumstances in order to explore a conceptual apparatus that they then apply in separate explanations, or just to rule out or rule in logical possibilities even before considering nomological or practical ones. There is much to learn about the nature of models in science independent of the metaphysical and epistemic issues that concern scientific realists and antirealists, and neither party to that dispute denies this.

A Case Study: Darwin's Theory of Natural Selection

More than once in Chapters 2 and 6, Darwin's theory of natural selection has been cited for its philosophical implications. In Chapter 2, it was argued that the theory encouraged some biologists to take on a number of questions that had hitherto been the preserve of philosophy: human nature, and even the existence of a divine creator whose design is reflected in biological phenomena. Chapter 6 raised the question of whether Darwin's account of adaptation as the result of purely causal processes provided an explanation of how purposes are possible in nature, or alternatively showed that there are no real purposes, goals, or ends in the biological realm and that all such descriptions were mistaken and illusory.

Darwin's achievement is sometimes treated as second only to Newton's in its revolutionary impact on science. For that reason, and because it is a theory from outside physics, it is worthwhile to employ it to illustrate and test the claims about theories and models that were made in the last three chapters. Moreover, the theory raises some philosophical problems that Chapter 10 will address more generally, concerning testability and confirmation.

In writing *On the Origin of Species*, Darwin did not lay out the theory of natural selection as a set of assumptions about an underlying mechanism from which a wide variety of generalizations about observable phenomena could be derived by deduction. And to this day, biologists, historians of science, and philosophers of science debate the exact structure of his theory. Some biologists and philosophers of science have been reluctant to extract a single set of laws of natural selection from the work, or from the sub-discipline of evolutionary biology that it spawned. These philosophers and biologists are not reluctant to expound the theory by offering a range of examples of how it works. Such examples are an effective way of introducing the theory.

Consider the Darwinian explanation for why all normal giraffes living today have long necks. Like all inherited traits, there is always variation in the length of giraffes' necks. At some time in the distant past, owing to random chance, a particularly long-necked variant appeared among a small number of giraffes. There is always variation—either mutation or genetic recombination—independent of, uncorrelated with, changes in the environment. This was one of Darwin's great discoveries. This small number of longer-necked giraffes did better at feeding than shorter-necked ones, and than other mammals competing with giraffes for resources, and so survived longer and had more longer-necked offspring, since neck-length is largely hereditary. Since the total giraffe population supported by its environment was limited, the proportion of longer-necked giraffes in the whole population increased as they out-competed the shorter-necked ones for limited resources (leaves high enough up on trees that only giraffes could reach them). Result: long necks eventually become uniform across the entire giraffe population.

Darwin presented a more general version of his theory in the following way. He began with two observations:

1.   Reproducing populations increase exponentially.
2.   The capacity of any region to support any reproducing population is finite.

Therefore, he inferred,

3.   There will always be a struggle for survival and reproduction among competing populations.

It is also obvious to observation that

4.   There is variation in the fitness of members of these populations and some of these variations are heritable.

Accordingly, Darwin concluded,

5.   In the struggle for survival and reproduction, the fittest variants will be favored and, therefore,
6.   Adaptive evolution will occur.

We might think of the three observations, 1, 2, and 4 as axioms, while treating 3, 5, and 6 as theorems.

Statement 1 is an observation that Darwin attributed to Thomas Malthus, a nineteenth-century economist who held that human populations increase geometrically while the food supply increases only arithmetically, and therefore "the poor will always be with us." The line of reasoning from this insight to Darwin's theory was in retrospect not hard to locate, even though the theory of natural selection is a much more consequential achievement.

It is statement 4 that expresses Darwin's distinctive discovery, that in the biological realm, unlike the chemical for example, variation is the norm. Compare the chemical elements: Among the purified samples of any element, all the features are the same. This is because at the level of the individual atom, every atom of the same element is the same, with the occasional exception of an isotope, which differs by a neutron that has no bearing on the atom's chemical reactions. By contrast, every member of a species is different from every other member in some way or other; even identical twins do not have exactly the same DNA sequences owing to copying errors in DNA replication. Darwin recognized that variation is "blind"—it is never the result of any sort of need or usefulness it might have for the organism. There is, he realized, no foresight in nature, driving variations. Darwin described the role of the environment as "natural selection." As a metaphor "natural selection" is inapt. For the environment never selects. Its role is purely passive. It acts as a filter, removing the less fit among members of reproductive lineages competing with one another and with members of other lineages. The environment, on Darwin's view, does not create adaptations; it doesn't even actively shape them. It merely lets through those variations minimally fit enough to survive another round of filtration as reproducers. Nature is not an active selector of novel variations it calls forth and chooses between.

In order to capture the theory of natural selection's generality, we can't express it as one about giraffes, or mammals, or animals, or even organisms. That is because as a general claim about a mechanism of evolution that could obtain anywhere in the universe at any time (something needed to make it a set of scientific laws), it can't mention things that are specific to the Earth. We need to express it as a claim about reproducing members of any line of (reproductive) descent. So stated the theory is not to be construed as a claim only about the evolution of plant and animal life on Earth. What is more, the lineages of reproducing members on the Earth include much more than the animals and plants: it will include genes, genomes (sets of genes on the same chromosome, for instance), single-celled asexual organisms, families, groups and populations, along with individual organisms—animals and plants. All these items reproduce, show heritable traits and variation in them, and so will participate in distinct evolutionary processes leading to adaptations at different levels of biological organization. Just as having long necks is an adaptation in giraffes whose distribution the theory explains, being able to survive in boiling water is an adaptation for certain bacteria, which enables the theory to explain their persistence in hot springs all over the world. In fact, every adaptation that biology has uncovered, no matter how intricate and complex, has its source in a wholly blind, purely passive process of random variation and environmental filtration.

Now we can see why some natural scientists, and some philosophers of science, have argued that being a purely causal theory, which has no room for purpose and teleology, Darwin's theory has overturned Kant's dictum that there will never be a Newton for a blade of grass. If they are correct, Darwin's mechanism of

blind variation and natural selection, along with its twentieth-century extensions (which explain heredity and variation in purely physical and chemical terms), represents a vindication of the scientific program of mechanism that began with Newton. It shows that the purely mechanical vision of nature associated with Newton's theory can be extended all the way through the life sciences, leaving no room for a teleological or purposive view of any part of nature.

Notice that, like other sets of laws, the theory of natural selection makes a number of hypothetical claims: If there is variation in heritable traits and if these variants differ in fitness, then there will be adaptational change. Like the kinetic theory of gases that tells us how gases behave, if they exist, without telling us that there are gases, Darwin's general theory does not assert that adaptational evolution obtains. For that conclusion we need initial conditions: the assertion that some things reproduce, that their offspring's traits are inherited from their parents, and that these traits are not always exact copies, but do in fact vary from parent to offspring and among offspring. *On the Origin of Species* of course made such assertions about lineages of the many plants and animals Darwin had been studying for 30 years by the time it was published in 1859. Like most other works of biology it describes a great deal about evolution on this particular planet, along with a general theory about evolution that could be realized by things elsewhere in the universe that look nothing like what we recognize as animals and plants, just so long as they show heritable variations in fitness to their environments.

Another thing to notice about Darwin's theory is that while evolution by natural selection requires reproduction with heritable variation, it is silent on how reproduction takes place, and tells us nothing about the mechanism of heredity. It presupposes that there is a mechanism of heredity, but it is silent on genetics, which is the mechanism of hereditary transmission on Earth. And of course, it must also be silent on the source of the variations that are continually being manifested from generation to generation, and among which the environment "selects" by filtering the less fit. Much twentieth-century biology has been devoted to providing the theory of how hereditary variation occurs on Earth that is required to apply Darwin's theory of natural selection in detail to explain the direction and rate of evolution on this planet over the last 3.5 billion years. This theory is of course molecular and population genetics.

Models and Theories in Evolutionary Biology

Unlike Newton, Darwin did not identify the central features of his theory as laws, nor did he express the process of blind variation and environmental filtration he discovered in a set of universal generalizations. He employed the word "laws" often enough in *On the Origin of Species* but not to refer to the mode of operation of the forces of evolutionary adaptation he discovered. Moreover, there are a number of other ways of stating or expressing his

theory besides Darwin's own. Some of them have the economy and simplicity of a Newtonian-style axiomatization. The trouble is that each of these presentations turns out to be, to greater or lesser degrees, inadequate to capture one or another of the processes that evolutionary biologists would describe as Darwinian evolution.

One of the most attractive of these presentations of the core assumptions, axioms, or underived laws of the theory of natural selection is due to Richard Lewontin. His presentation of the theory takes the form of three claims, each of which is necessary and all taken together are sufficient for the evolution of adaptations:

1. There is always variation in the traits of whatever it is that replicates or reproduces.
2. The variant traits differ in fitness.
3. The fitness differences among some of the traits are heritable.

Notice that this presentation is silent on the blindness of variations or the passive role of the environment that determines fitness differences. Even if this expression of the theory is adequate (and this has been disputed in several different ways), it is far too abstract to be applied to explain evolutionary processes or outcomes, and its key concepts—traits, fitness, replication/reproduction, heritability—are open to interpretations that make the theory false in some cases, and irrelevant to evolution in other cases.

This of course is just what one might expect on the semantic approach that treats the theory as consisting in a large number of models not all of which apply to all cases. The three conditions that Lewontin articulated are too abstract to be anything but "schemata"—very loose and open frameworks that need to be filled in to have any content, but which can be filled in to make models that illustrate or realize these principles as a theory, in the way that semantic theory advocates. They can be filled in and then harnessed together with the details of various biological processes to explain how they emerged, and why they persist over time.

Some of the most famous insights of evolutionary biology are models that are driven by natural selection, but in which its role is not apparent. Almost any model that explains current distributions of flora and fauna as the result of a local or global equilibrium balancing selective forces illustrates the centrality of models in evolutionary biology. Among the most obvious are the two simple Lotka–Volterra equations that model the cyclic relations between populations of predators and prey, parasites and parasitized:

$$dx / dt = x(a - by)$$
$$dy / dt = -y(g - dx)$$

where y is the prey population number and x is the predator population number, dy/dt and dx/dt are the growth rates of the two populations, and a, b, g and d are parameters representing the interaction rates of the two populations. As predators increase in number, prey will be reduced with a certain lag expressed by the equations. When prey populations fall, predator populations will fall too, again with a lag, so that both populations cycle around equilibrium values. Applying the two equations requires a good deal of empirical research to estimate the length of the lag, and the actual population levels, which will differ for every pair of predators and prey. And of course there are cases where the equilibrium breaks down, for well-understood reasons. These are domains in which the model doesn't work. They have no tendency to undermine the model's explanatory or predictive role in other domains of application, and no one in ecology is searching for a more general set of laws to which these models approximate, just as the semantic approach would suggest.

Another example of this sort of equilibrium modeling is even more famous and characteristic of the centrality of models in evolutionary biology. The "Fisher sex-ratio model" explains why the ratio of males to females in almost all sexually reproducing species is conveniently 1 to 1—just enough of each so that everyone can expect one and only one partner. The model treats the female's disposition to have mainly male or mainly female offspring as a heritable trait. It uses the Lewontin assumptions to derive the conclusion that the former trait will be fitter in an environment disproportionately female, and vice versa. If fitness is a matter of having more offspring, then in the next generation females disposed to have more males will be fitter, and there will be more males, thus reducing the disproportion of females. This process continues until there are disproportionately too many males and then cycles back, always keeping the number of males and females in rough equality of numbers. The role of the model is to show that a well-established regularity that appears to be a convenient miracle is actually a natural consequence of Darwinian processes.

When biologists specify differing subject matters for the theory, sexual vs. asexual species, plants vs. animals, genes vs. individual organisms vs. families of individuals, with different mechanisms and rates of variation in hereditary transmission, they generate different models of evolution by natural selection. The generic (not genetic) statement of the theory is too abstract and has insufficient content on this view to count as the theory of natural selection that biologists will recognize. But the wide range of models has enough structure in common to constitute a family of models, just as the semantic theory suggests.

There is another powerful reason to find the semantic view of Darwinian theory attractive. The problem stems from what is perhaps the oldest and at the same time most vexing problem facing the theory of natural selection. It was a nineteenth-century philosopher, Herbert Spencer, who characterized Darwinism as the theory of "the survival of the fittest," meaning that the fittest will survive to out-reproduce the less fit and by iteration produce evolution. And the label

"survival of the fittest" has stuck. Indeed it is not inapt, for it appears that one central claim of the theory can be expressed as follows—the principle of natural selection (PNS):

> PNS: Given two competing populations, x and y, if x is fitter than y, then in the long run, x will leave more offspring than y.

The PNS is not expressed in the Lewontin formulation of the theory, but it is not difficult to see that it or something very like it is required. Heritable variations in fitness among reproducing entities won't lead to changes in the distribution of traits without making a difference in reproductive rates.

But trouble arises for the theory when we ask what "fitter than" means. If the PNS is to be a contingent empirical law, then one thing we have to rule out is that differences in fitness are defined as differences in number of offspring left in the long run. For that would turn the PNS into the explanatorily uninformative necessary truth that "if x leaves more offspring than y in the long run, then in the long run x will leave more offspring than y." Logically necessary truths cannot be scientific laws, and cannot explain any contingent empirical fact. The PNS could explain differences in offspring numbers on this meaning of fitness only if events (like having more offspring) can provide their own explanations—something we ruled out in Chapter 2.

We could of course refuse to define fitness. Instead we could just hold, along with realists about theoretical entities, that "fitness" is a theoretical term, like "positive charge" or "atomic mass." But that seems implausible and unsatisfying. After all, we know that taller giraffes and speedier zebras are fitter without the aid of instruments of indirect observation; we know what fitness is … it's the organism's ability to solve problems presented to it by the environment: avoiding predators, securing prey, keeping sufficiently warm and dry (unless a fish), etc. But why are these the problems that an organism must solve to be fit? How do they combine into overall fitness? How do we compare organisms for fitness when their abilities to solve any one of these problems differ? The most reasonable answers to these questions appear to be that (a) the problems that the environment presents an organism with are ones whose solution increases the organisms' chances to survive and reproduce, (b) we can calculate the degree to which organisms solve these various problems by measuring the organisms' number of offspring, and (c) two organisms are equally fit no matter how differently they deal with environmental problems, provided they have the same number of offspring. The only thing wrong with these answers is that they show how almost inevitable is the temptation to define "fitness" in terms of reproduction, thus turning the PNS itself into a definition.

The proponent of the semantic approach to theories has little difficulty with this outcome. The semantic theory can accept that the PNS is a definition; theories are made up of definitions like the PNS along with claims about the different things in the world that satisfy this definition. The variety of things on the Earth,

let alone on other worlds in other galaxies, that can realize or instantiate an evolutionary process (whether it be genes, organisms, groups or cultures) seems to cry out for a semantic approach to Darwinism. The theory's silence on the detailed mechanisms that provide the heredity and the variations in hereditary traits required for evolution here on Earth—nucleic acids and mutations in them—are presumably mechanisms quite different from what we can anticipate finding elsewhere in the universe. This is yet another reason to treat Darwinian theory as a set of models that can be realized in many different ways by many different systems.

Yet a problem remains for the semantic approach to the theory of natural selection. On the semantic approach a scientific theory is really more than the set of the models that take its name. It's that set along with the assertion that things in the world realize, satisfy, instantiate, and exemplify these definitions sufficiently well to enable us to explain and predict their behavior (observable or unobservable) to some degree of accuracy. Without this further assertion, a scientific theory would be no different from a piece of pure set-theory. So, even the proponent of the semantic theory must recognize that asserting a theory is to make a substantive claim about the world, in particular it is to say that the same causal process is at work making all these different phenomena satisfy the same definition. Thus, in the end, like the axiomatic account, the semantic approach is committed to the truth of some general claims that themselves cry out for explanation. It is not really enough to identify a set of models that share a structure in common and are applicable to a diversity of empirical phenomena, then not explain why they do so. Unless we find ourselves at the end of inquiry when no further explanations of the fundamental laws of nature can be given, there will have to be some underlying mechanism or process that is shared among all the different things that realize the same set-theoretical definition, an underlying mechanism that explains why the predictions we can make employing the model are confirmed. Thus, the semantic view of theories has all the same intellectual obligations to explain why theories are true or approximately true, or at least moving successively closer to the truth, than the axiomatic account does. That is, it is also committed to the truth of some substantive general laws about the way things are in the world, laws about natural selection among them. So, in the end, it will have to face the problems raised by the role that "fitness" plays as the key explanatory variable in Darwinian theory.

Summary

The axiomatic approach to theories has difficulty accommodating the role of models in science. Instrumentalism does not, and as models become more central to the character of scientific theorizing, problems for the axiomatic approach and for realism mount. The issue here ultimately turns on whether science shows a pattern of explanatory and predictive successes that can only be explained by realism and the truth of theories that organize and explain the success of the models that scientists develop.

Darwin's theory of natural selection provides a useful "test bed" for applying and assessing the adequacy of some of the competing conceptions of scientific theories articulated in this chapter. There are several reasons for thinking that Darwin's theory is quite unlike Newton's despite their very similar roles in organizing almost all of their disciplines. For one thing, it is hard to state any laws of Darwinian natural selection, or indeed any strict laws in biology at all. For another, almost all of the explanatory applications of the theory of natural selection proceed by the construction of models. And most of these models are not obviously judged on their predictive success, owing to the complexity of biological systems.

Both in biology and the special sciences, the role of models in advancing scientific understanding seems much more nuanced and diverse than the way that philosophers of science traditionally have pictured theories operating in physics. This has opened up a broad area of research in the philosophy of science.

Study Questions

1. What makes the semantic approach, with its emphasis on models, more amenable to instrumentalism than to realism?
2. Defend or criticize: "No matter what the motive for constructing them, all models are ultimately to be judged on their roles in prediction."
3. Does physics differ from biology in its employment of models just because general theories seem to be available in that discipline?
4. A famous biologist argued that there are unavoidable trade-offs in modeling between realism, generality, and precision. Is this true for models in all disciplines, or just in biology?
5. Can the causal mechanism of variation and selection that Darwin uncovered be applied to explain the purposive character of phenomena beyond those of interest strictly to biologists, such as anatomy? For example, can it be employed to explain human behaviors and human social intuitions as the results of variation and environmental selection, and not the conscious choice of individuals or groups of them?

Suggested Readings

The semantic view of theories is elaborated by F. Suppe in *The Structure of Scientific Theories*, as well as by van Fraassen, *The Scientific Image*. Ronald Giere develops a model-based account of science in *Explaining Science* and *Science without Laws*. Its application to biology is treated in P. Thompson, *The Structure of Biological Theories*, and E. Lloyd, *The Structure of Evolutionary Theory*.

M. S. Morgan and M. Morrison, *Models as Mediators: Perspectives on Natural and Social Science*, provides still another account of models and their role in

science, as representations, tools, test beds that are not judged on exclusively explanatory and predictive criteria.

There is no substitute for reading *On the Origin of Species*, but Dawkins' *The Blind Watchmaker* comes close. General introductions to the nature of Darwin's theory are to be found in E. Sober, *Philosophy of Biology* and *The Nature of Selection*, as well as Daniel McShea and A. Rosenberg, *Philosophy of Biology: A Contemporary Approach*, and K. Sterelny and P. Griffiths, *Sex and Death*.

Lange's anthology reprints a powerful argument for the absence of strict laws in biology, John Beatty, "The Evolutionary Contingency Thesis."

10 Induction and Probability

Overview

Suppose we settle the dispute between realism and instrumentalism. The problem still remains of how exactly observation and evidence, the collection of data, etc., actually enable us to choose among scientific theories. On the one hand, that they do so has been taken for granted across several centuries of science and its philosophy. On the other hand, no one has fully explained how they do so, and in this century the challenges facing the explanation of exactly how evidence controls theory have increased.

A brief review of the history of British empiricism sets the agenda for an account of how science produces knowledge justified by experience, and introduces the problem of induction raised by David Hume in the eighteenth century. If we cannot solve the problem of induction, we may be able to show that it is a pseudo-problem. Even if we can't do that, scientists are not about drop their tools and wait for a resolution of this matter. What is more, they may insist that they know how to proceed inductively without any help from philosophers. All we really need, many scientists would insist, is a theorem of the probability calculus derived in the eighteenth century by Thomas Bayes, a contemporary of David Hume. Some philosophers will concur with this judgment. So we need to understand this theorem and the interpretational issues that are raised by its use in experimental and observational reasoning.

The Problem of Induction

As we noted in Chapter 7, the scientific revolution began in central Europe with Copernicus, Brahe and Kepler, shifted to Galileo's Italy, moved to Descartes' France, and came to Newton in Cambridge, England. The scientific revolution was also a philosophical revolution, for reasons we noted. In the seventeenth century science was "natural philosophy," and figures that history would consign exclusively to one or the other of these two fields—philosophy or science—contributed to both. Thus Newton wrote a good deal of philosophy of science, and Descartes made contributions to physics. But it was the British empiricists who made a self-conscious attempt to examine whether the theory of knowledge espoused by these scientists would vindicate the methods that Newton, Boyle, Harvey, and other experimental scientists employed to expand the frontiers of human knowledge so vastly in their time.

Over a period from the late seventeenth century to the late eighteenth century, John Locke, George Berkeley, and David Hume sought to specify the nature, extent and justification of knowledge as founded on sensory experience and to consider whether it would certify the scientific discoveries of their time as knowledge and insulate them against skepticism. Their results were mixed, as Kant was eager to point out. But nothing would shake their confidence, or that of most scientists, in empiricism as the right epistemology.

Locke sought to develop empiricism about knowledge, famously holding against rationalists like Descartes, that there are no innate ideas. "Nothing is in the mind that was not first in the senses." But Locke was resolutely a scientific realist about the theoretical entities that seventeenth-century science was uncovering. As noted in Chapter 2, he embraced the view that matter was composed of indiscernible atoms, or "corpuscles," and distinguished between material substance and its properties ("primary qualities") on the one hand, and the sensory qualities of color, texture, smell or taste (the so-called secondary qualities), that substances cause in us. The real properties of matter, according to Locke, are just the ones that Newtonian mechanics tells us it has—mass, extension in space, velocity, etc. The sensory qualities of things are ideas in our heads that the things cause. It is by reasoning back from sensory effects to physical causes that we acquire knowledge of the world, which gets systematized by science.

That Locke's realism and empiricism inevitably gives rise to skepticism is not something Locke recognized. It was a philosopher of the next generation, George Berkeley, who appreciated that empiricism makes doubtful our beliefs about things that we do not directly observe. How could Locke lay claim to certain knowledge of the existence of matter or its features, if he could only be aware of sensory qualities which, by their very nature, exist only in the mind? We cannot compare sensory features like color or texture to their causes to see whether these causes are colorless or not, for we have no access to these things, and so cannot compare them. And to the argument that we can imagine something to be colorless, but we cannot imagine a material object to lack extension

or mass, Berkeley retorted that sensory properties and non-sensory ones are on a par in this respect: try to imagine something without color. If you think of it as transparent, then you are adding in the background color and that's cheating. Similarly for the other allegedly subjective qualities that things cause us to experience.

In Berkeley's view, without empiricism we cannot make sense of the meaningfulness of language. Berkeley pretty much adopted the theory of language as naming sensory qualities that was sketched in Chapter 8. Given the thesis that words name sensory ideas, realism—the thesis that science discovers the truth about things that we cannot have sensory experience of, becomes false, for the words that name these things must be meaningless. In place of realism Berkeley advocated a strong form of instrumentalism and took great pains to construct an interpretation of seventeenth- and eighteenth-century science, including Newtonian mechanics, as a body of heuristic devices, calculating rules, and convenient fictions that we employ to organize our experiences. Doing this, Berkeley thought, saves science from skepticism. It did not occur to him that another alternative to the combination of empiricism and instrumentalism is rationalism and realism. And the reason is that by the eighteenth century, the role of experiment in science had been so securely established that no alternative to empiricism seemed remotely plausible as an epistemology for science. Even rationalism, as we noted in Chapter 1, argues only that some scientific knowledge has a non-empirical justification.

It was David Hume's intention to apply what he took to be the empirical methods of scientific inquiry to philosophy. Like Locke and Berkeley he sought to show how knowledge, and especially scientific knowledge, honors the strictures of empiricism. Unable to adopt Berkeley's radical instrumentalism, Hume sought to explain why we adopt a realistic interpretation of science and ordinary beliefs, without taking sides between realism and instrumentalism. But Hume's pursuit of the program of empiricism led him to face a problem different from that raised by the conflict of realism and empiricism. This is the problem of induction: Given our current sensory experience, how can we justify inferences from them and from our records of the past, to the future and to the sorts of scientific laws and theories that we seek?

Hume's argument is often reconstructed as follows: There are two and only two ways to justify a conclusion: deductive argument, in which the conclusion follows logically from the premises, and inductive argument, in which the premises support the conclusion but do not guarantee it. A deductive argument is colloquially described as one in which the premises "contain" the conclusion, whereas an inductive argument is often described as one that moves from the particular to the general, as when we infer from observation of 100 white swans to the conclusion that all swans are white. Now, if we are challenged to justify the claim that inductive arguments—arguments from the particular to the general, or from the past to the future—will be reliable in the future, we can do so only by employing a deductive argument or an inductive argument. The trouble with any deductive argument to this conclusion is that at least one of the premises will

itself require the reliability of induction. For example, consider the deductive argument below:

1. If a practice has been reliable in the past, it will be reliable in the future.
2. In the past inductive arguments have been reliable.
      Therefore,
3. Inductive arguments will be reliable in the future.

This argument is deductively valid, but its first premise requires justification and the only satisfactory justification for the premise would be the reliability of induction, which is what the argument is supposed to establish. Any deductive argument for the reliability of induction will therefore include at least one question-begging premise.

This leaves only inductive arguments to justify induction. But clearly, no inductive argument for induction will support its reliability, for such arguments too are question-begging. As we have had occasion to note before, like all such question-begging arguments, an inductive argument for the reliability of induction is like underwriting your promise to pay back a loan by promising that you will keep your promises. If your reliability as a promise keeper is what is in question, offering a second promise to assure the first one is pointless. Hume's argument has for 250 years been treated as an argument for skepticism about empirical science. It suggests that all conclusions about scientific laws, and all predictions that science makes about future events, are at bottom unwarranted, owing to their reliance on induction. And it's not just inferences from the specific to the general or from the past to the future. There are other forms of argument that are clearly inductive without taking either of these forms, including arguments by analogy, and inferences to the best explanation as employed to infer the existence of unobservable entities throughout the sciences. All ampliative forms of argument, in which the conclusions are intended to make claims that transcend those of the premises, will be inductive and open to Hume's challenge. Many ampliative inferences employ or exploit deduction, but they are inductive nevertheless. For example, hypothetico-deductive reasoning involves the deduction of observational consequences from a hypothesis but their direct testing is still inductive. If confirmed, such deductive consequences are said to confirm the hypothesis that they are deduced from. The whole inference is clearly inductive: The conclusion, that the credence of the general hypothesis is strengthened by the narrower observational evidence, goes beyond the evidence.

Hume's challenge is theoretical. He noted that as a person who acts in the world, he was satisfied that inductive arguments were reasonable; what he thought the argument shows is that we have not yet found the right justification for induction, not that there is no justification for it.

Hume's problem of induction was surprisingly invisible for the first 150 years after he formulated it. The greatest empiricist epistemologist and philosopher of science of the nineteenth century, John Stuart Mill, completely failed to recognize it despite devoting much attention to induction as the core method of

science. According to Mill, inferences from a relatively small number of cases to general laws are how science proceeds. Mill famously articulated several rules of experimental design that still guide scientists today in making such inferences. Contemporary double-blind controlled experiments, now commonplace in medical science, owe a great deal to the rules that Mill set out and the arguments he gave for them.

But whether the practice of inductive inference required independent justification as a whole was something Mill did not appreciate. Mill believed with some justification that inductive inferences were grounded on a commitment to the uniformity of nature: that the future will be like the past. If we can be justified in believing this principle then at least some inductive inferences will be warranted. But what sort of argument can be advanced for the uniformity of nature? A deductive argument with a factual conclusion that the future will be like the past will have to include among its premises a factual claim at least as strong, and this will then require justification, and so on in an infinite regress. An inductive argument for the uniformity of nature will proceed along the following lines: In the recent past, its near future was like the more distant past, in the more distant past, its near future was like the even more distant past, and so on. Therefore, hereafter the future will be like the recent past, the more distant past, and the very distant past. But this form of argument is itself inductive and so begs the question. We set out to establish the reliability of inductive inference and do so by an inductive inference. As we have seen, this has all the reliability of an attempt to assure someone that I keep my promises by promising him that I do so!

During the period in which the logical positivists were confident that the principles of mathematical and symbolic logic were definitions and consequences of them, attempts were made to solve Hume's problem in a similar way. Philosophers like Rudolph Carnap and Carl G. Hempel sought to frame rules of inductive inference that could be justified, like the laws of mathematical logic, on the basis of definitions and their implications. Like the deductive-nomological model that was proposed to rationally reconstruct the concept of scientific explanation, their aim was to provide a "confirmation theory" that would formalize and explicate the notion of inductive inference and solve Hume's problem too. The strategy was to show that inductive argument turns out to be deductive argument that employs special rules that confer justification on their conclusions without guaranteeing their truth (unlike the laws of deductive logic which did so). These rules would reflect the axioms and theorems of probability theory, a set of logical truths or definitions. In order for these rules to systematize inductive inferences, the statements that scientists use to describe the data or evidence to which the rules are applied had to be given a rigid logical structure and a wholly observational vocabulary. This could not reasonably accommodate the actual patterns of scientific inference. But in addition, the entire enterprise of developing a purely formal or logical theory of probability only revealed the problem to be even more serious than Hume had recognized, as we will see in Chapter 11.

Other philosophers sought to show that the problem of induction was a pseudo-problem, a classical example of the bewitchment of our understanding by language. Thus it has been repeatedly argued that to employ inductive principles in order to frame expectations about the future is just common sense and what most people mean by being reasonable. If employing inductive inference is, by definition, a necessary condition for acting in a reasonable manner, then it is senseless to demand a justification for it. Or at least it makes no more sense to ask that induction be shown to be reasonable than it makes to ask that being reasonable be shown to be reasonable. Thus, a proper understanding of what it means to be reasonable when framing beliefs about the unobserved solves the problem of induction, or rather shows it to be a pseudo-problem, reflecting mistakes about language. What mistake? One candidate is the tendency mistakenly to apply deductive standards to induction and then to complain when they can't be met. Validity is a feature of proper deductive arguments: these arguments are always truth-preserving. Since inductive arguments are by their nature *not* truth-preserving (nor intended or expected to be), it is easy but mistaken to describe them as invalid and then demand a justification for them. The mistake is even to apply the valid/invalid distinction to such arguments and then to demand a substitute for validity.

Few philosophers of science could take seriously this way of dismissing Hume's problem. They insisted that the callow mistake identified by those who seek to dissolve the problem of induction is not one they make. The problem of induction is very clearly that of showing inductive inferences to be generally reliable, not universally valid. And this problem can be framed in such a way as to honor the thought that being inductive is being reasonable. To ask whether being reasonable, i.e. using inductive methods, is a reliable method of getting through life is perfectly intelligible. The question of whether being reasonable is reliable is one that we all want to answer affirmatively. Hume in effect invites us to do so in a non-question-begging way.

One way to respond to Hume that recognizes this way of putting his problem was due to the logical positivist philosopher Hans Reichenbach (he preferred the label "logical empiricist"). He sought to show that if any method of predicting the future works, then induction must work. Suppose we wish to establish whether the Oracle at Delphi is an accurate predictive device. The only way to do so is to subject the Oracle to a set of tests: ask for a series of predictions and determine whether they are verified. If they are, the Oracle can be accepted as an accurate predictor. If not, then the future accuracy of the Oracle is not to be relied upon. But notice that the form of this argument is inductive: if any method works (in the past), only induction can tell us that it does (in the future). Whence we secure the "vindication" of induction. This argument faces two difficulties. First, at most it proves that if any method works, induction works. But this is a far cry from the conclusion we want: that there is any method that does in fact work. Second, the argument will not sway the devotee of the Oracle. Oracle-believers will have no reason to accept our argument. They will ask the Oracle whether induction works, and will accept its pronouncement. No attempt to convince

Oracle-believers that induction supports either their method of telling the future or any other can carry any weight with them. The argument that if any method works, induction works, is question-begging too.

Statistics and Probability to the Rescue?

At some point the problems of induction will lead some scientists and philosophers to lose patience with the philosophy of science. Why worry about justifying induction? Why not get on with the serious but perhaps more soluble problem of empirical confirmation? We may grant the fallibility of science, the impossibility of establishing the truth or falsity of scientific laws once and for all. Yet we may still explain how observation, data collection, and experiment test scientific theories by turning to statistical theory and the notion of probability.

Yet it turns out that doing so is not as simple a matter as it may seem. To begin with, the notions of probability and empirical or inductive evidence don't really line up together as neatly as we might wish.

First, there is the problem of whether the fact that some data raises the probability of a hypothesis makes that data evidence for it at all. This may sound like a question trivially easy to answer, but it isn't. Define $p(h, b)$ as the probability of hypothesis h, given background information b, and $p(h, e$ and $b)$ as the probability of h given the background information b, and some experimental observations e. Suppose we adopt the principle that

e is positive evidence for hypothesis h if and only if $p(h, e$ and $b) > p(h, b)$

So, when data increases the probability of a hypothesis, it constitutes favorable evidence for it.

In this case, e is "new" data that counts as evidence for h if it raises the probability of h (given the background information required to test h). For example, the probability that the butler did it, h, given that the gun found at the body was not his, b, but the new evidence that the gun carried his fingerprints, e, is higher than the hypothesis that the butler did it, given the gun found at the body, and no evidence about fingerprints. It is the fingerprints that raise the probability of h. That's why the prints are evidence that the butler did it.

It is easy to construct counterexamples to this definition of positive evidence that show that increasing probability is by itself neither necessary nor sufficient for some statement about observations to confirm a hypothesis. Here are two:

This book's publication increases the probability that it will be turned into a blockbuster film starring Keira Knightley. After all, were it never to have been published the chances of its being made into a film would be even smaller than they are. But surely, the actual publication of this book is not positive evidence for the hypothesis that this book will be turned into a blockbuster film starring Keira Knightley. It is certainly not clear that some fact that just raises the probability of a hypothesis thereby constitutes positive evidence for it. A similar

conclusion can be derived from the following counterexample, which invokes lotteries, a useful notion when exploring issues about probability. Consider a fair lottery with 1,000 tickets, 10 of which are purchased by Andy and 1 of which is purchased by Betty. h is the hypothesis that Betty wins the lottery. e is the observation that all tickets except those of Andy and Betty are destroyed before the drawing. e certainly increases the probability of h from 0.001 to 0.1. But it is not clear that e is positive evidence that h is true. In fact, it seems more reasonable to say that e is positive evidence that h is untrue, that Andy will win. For the probability that he wins has gone from 0.01 to 0.9. Another lottery case suggests that raising probability is not necessary for being positive evidence; indeed a piece of positive evidence may lower the probability of the hypothesis it confirms. Suppose in our lottery Andy has purchased 999 tickets out of 1,000 sold on Monday. Suppose e is the evidence that by Tuesday 1,001 tickets have been sold of which Andy purchased 999. This e lowers the probability that Andy will win the lottery from 0.999 to 0.998 … But surely e is still evidence that Andy will win after all.

One way to deal with these two counterexamples is simply to require that e is positive evidence for h if e makes h's probability high, say above 0.5. Then in the first case, since the evidence doesn't raise the probability of Betty's winning anywhere near 0.5, and in the first case the evidence does not lower the probability of Andy's winning much below 0.999, these cases don't undermine the definition of positive evidence when so revised. But of course, it is easy to construct a counterexample to this new definition of positive evidence as evidence that makes the hypothesis highly probable. Here is a famous case: h is the hypotheses that Andy is not pregnant, while e is the statement that Andy eats Weetabix breakfast cereal. Since the probability of h is extremely high, p(h, e)—the probability of h, given e—is also extremely high. Yet e is certainly no evidence for h. Of course we have neglected the background information, b, built into the definition. Surely if we add the background information that no man has ever become pregnant, then p(h, e & b)—the probability of h, given e and b—will be the same as p(h, e), and thus dispose of the counterexample. But if b is the statement that no man has ever become pregnant, and e is the statement that Andy ate Weetabix, and h is the statement that Andy is not pregnant, then p(h, e & b) will be very high, indeed about as close to 1 as a probability can get. So, even though e is not by itself positive evidence for h, e plus b is, just because b is positive evidence for h. We cannot exclude e as positive evidence, when e plus b is evidence, just because it is a conjunct that by itself has no impact on the probability of h, because sometimes positive evidence does only raise the probability of a hypothesis when it is combined with other data. Of course we want to say that in this case, e could be eliminated without reducing the probability of h. E is probabilistically irrelevant and that's why it is not positive evidence. But providing a litmus test for probabilistic irrelevance is no easy task. It may be as difficult as defining a positive instance. In any case, we have an introduction here to the difficulties of expounding the notion of evidence in terms of the concept of probability.

Philosophers of science who insist that probability theory suffices to enable us to understand how data test hypotheses will respond to these problems that they reflect the misfit between probability and our common-sense notions of evidence. Our ordinary concepts are qualitative, imprecise, and not the result of a careful study of their implications. Probability is a quantitative mathematical notion with secure logical foundations. That enables us to make distinctions that ordinary notions cannot draw, and to explain these distinctions. Recall the logical empiricists, who sought rational reconstruction or explication of concepts like explanation that provide necessary and sufficient conditions in place of the imprecision and vagueness of ordinary language. Likewise, many contemporary students of the problem of confirmation seek a more precise substitute for the ordinary notion of evidence in the quantifiable notion of probability; for them counterexamples such as the ones above simply reflect the fact that the two concepts are not identical. They are no reason not to substitute "probability" for "evidence" in our inquiry about how data test theory. Some of these philosophers go further and argue that there is no such thing as evidence confirming or disconfirming a hypothesis by itself. Hypothesis testing in science is always a comparative affair: It only makes sense to say hypothesis h1 is more or less well confirmed by the evidence than is hypothesis h2, not that h1 is confirmed by e in any absolute sense.

These philosophers hold that the mathematical theory of probability holds the key to understanding the confirmation of scientific theory. And this mathematical theory is extremely simple. It embodies only three very obvious assumptions:

1. Probabilities are measured in numbers from 0 to 1.
2. The probability of a necessary truth (like "4 is an even number") is 1.
3. If hypothesis h and j are incompatible, then p(h or j) = p(h) + p(j).

It's easy to illustrate these axioms with a deck of normal playing cards. The probability of any one card being drawn from a complete deck is between 0 and 1. In fact it's 1/52. The probability that a card will be red or black (the only two possibilities) is 1 (it's a certainty), and if drawing an ace of hearts is incompatible with drawing a jack of spades, then the probability of drawing one of them is 1/52 + 1/52, or 1/26, about 0.038461…

From these simple and straightforward assumptions (plus some definitions), the rest of the mathematical theory of probability can be derived by logical deduction alone. In particular, from these three axioms of the theory of probability, we can derive a theorem, first proved by a British theologian and amateur mathematician in the eighteenth century, Thomas Bayes, which has loomed large in contemporary discussions of confirmation. Before introducing this theorem, we need to define one more notion, the conditional probability of any one statement, assuming the truth of another statement. The conditional probability of a hypothesis, h, on a description of data, e, written p(h/e) is defined as the ratio of the probability of the truth of both h and e to the probability of the truth of e alone:
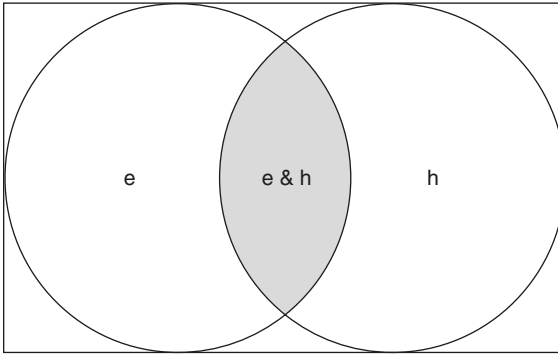
Figure 10.1 Circles e and h are the same size, and between them take up most of the rect-
angle, to suggest that the probability of a dart hitting one of them (and not
the other) is large and about the same.

$$p(h/e) = \frac{df\ p(h\ and\ e)}{p(e)}$$

Roughly, "the conditional probability of h on e" measures the proportion of
the probability that e is true which "contains" the probability that h is also true.
Adapting an expository idea of Martin Curd and Jan Cover, we can illuminate
this definition with a few diagrams. Suppose we are throwing darts at a board
on which two overlapping circles are drawn in the shape of a Venn diagram
(Figure 10.1).

If a dart lands inside circle e, what is the probability that it will also land inside
circle h, i.e. the probability of landing in h, on the condition that it lands in e, the
conditional probability, p(h/e)? That depends on two things: the area of overlap
between circle e and circle h (the intersection e & h), relative to the area of e, and
the size of e compared to the size of h. To see this, compare the two following
diagrams. In Figure 10.2, e is very large compared to the size of h, so the chance
that a dart thrown inside e also lands in h is low. But it would be higher if more
of h were inside e. On the other hand, the chance that a dart that lands in h also
lands in e is much higher, and increases as the proportion of h inside e grows.

By contrast, consider Figure 10.3. Here e is small and h is large. In this case the
chance of a dart that lands in e also landing in h is higher than in the previous
case, and becomes even higher the more of e is inside h. Again, the conditional
probability of e on h is of course much lower, the smaller the h circle is and the
less it overlaps.

The definition of conditional probability incorporates these two factors on
which conditional probability depends. The numerator reflects the size of the
overlap of e and h relative to the sizes of e and h, and the denominator measures
that size in units of e's size.
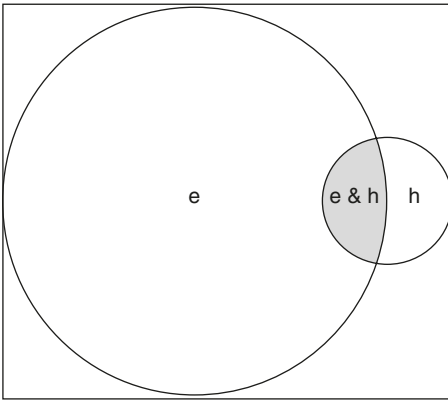
Figure 10.2 Circle e is much larger than circle h, so the probability of a dart hitting e is much higher than the probability of a dart hitting h. The shaded intersection e & h is much smaller than e, and a relatively large proportion of h. Thus p(h/e) is low, and p(e/h) is much higher than p(h/e).
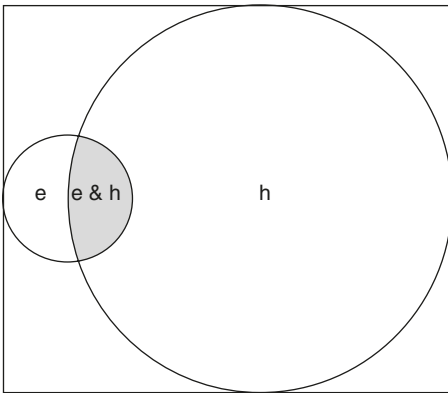
Figure 10.3 Circle h is much larger than circle e, so the probability of a dart hitting h is much higher than the probability of a dart hitting e. The shaded intersection e & h is much smaller than h, and is a relatively large proportion of e. Thus p(h/e) is high, and p(e/h) is much lower than p(h/e).

Now if h is a hypothesis and e is a report of data, Bayes' theorem allows us to calculate the conditional probability of h on e, p(h/e). In other words, Bayes' theorem give us a mathematical formula for calculating how much more or less probable a bit of evidence, e, makes any hypothesis, h. The formula is as follows:

$$p(h/e) = \frac{p(e/h) \times p(h)}{p(e)}$$

Bayes' theorem tells us that once we acquire some data, e, we can calculate how the data e changes the probability of h, raising it or lowering it, provided we already have three other numbers:

p(e/h)—the probability that e is true assuming that h is true (as noted above, not to be confused with p(h/e), the probability that h is true, given e, which is what we are calculating). This number reflects the degree to which our hypotheses leads us to expect the data we have gathered. If the data is just what the hypothesis predicts, then of course p(e/h) is very high. If the data is nothing like what the hypothesis predicts, p(e/h) is low.

p(h)—the probability of the hypothesis independent of the test which the data described by e provides. If e reports new experimental data, then p(h) is just the probability the scientist assigned to h before the experiment was conducted.

p(e)—the probability that the statement describing the data is true independent of whether h is true or not. Where e is a surprising result which previous scientific theory and evidence (independent of h) does not lead us to expect, p(e) will be low.

To see how easily Bayes' theorem follows from the axioms of probability and our definition of conditional probability, return to any of the dart-board diagrams above. If we can calculate p(e/h) by comparing the relative sizes of the circles and the ratio of their intersections to their sizes, we can also calculate p(h/e) the same way. Of course the figures for each conditional probability will be different (as each of the diagrams illustrates).

By drawing e- and h-circles and intersections of them of different sizes, it is easy to see that the probability of a dart that hits the e-circle also hitting the h-circle, p(h/e) will vary directly as the ratio of the intersection of the two circles to the size of the e-circle, and inversely as the ratio of the sizes of e-circle to the size of the h-circle. And this is exactly what Bayes' theorem says: it makes p(h/e) equal to p(e/h)—the ratio of the intersection of e and h to the size of e—times the fraction p(h)/p(e), which is the ratio of the size of h to the size of e.

Two simple examples may help us to see how Bayes' theorem is supposed to work: Consider how data on the observed position of Halley's comet provide a test for Newton's laws. Suppose, given prior observations, that p(e), the probability that Halley's comet will be observed in a particular location of the night sky, is 0.8. This allows for imperfections in the telescope, atmospheric irregularities, all the factors that eventually led astronomers to take many photographs of the stars and planets and to average their positions to make estimates of their expected positions in the sky. p(e/h) is also high, the expected position of Halley's comet in the night sky is very close to what the theory predicts it would be. Let's set p(e/h) at 0.95. Let's assume that prior to the acquisition of e, the new data about Halley's comet, the probability that Newton's laws are true is, say, 0.8. Thus, if Halley's comet appears where expected, p(h/e) = (0.95…) × (0.8)/

(0.8) = 0.95. Thus, the evidence as described by e has raised the probability of Newton's laws from 0.8 to 0.95.

But now, suppose we acquire new data about, say, the precession of the perihelion of Mercury—that is, data that show that the elliptical orbit of Mercury around the Sun is itself swinging so that the closest point between Mercury and the Sun keeps shifting. Suppose, as was indeed the case, that the figure turns out to be much higher than Newton's laws (and the auxiliary hypotheses used to apply them) would lead us to expect, so that p(e/h) is low, say 0.3. Since Newton's laws did not lead us to expect this data, the **prior probability** of e must be low, so let's let p(e) be low, say, 0.2; and the prior probability of such unexpected data, given Newton's laws plus auxiliary hypotheses, will also be quite low, say, p(e/h) is 0.1. If p(h) for Newton's laws plus auxiliaries is 0.95, then Bayes' theorem tells us that for the new e, the precession data for Mercury, the p(h/e) = (0.1) × (0.95)/ (0.2) = 0.475, a significant drop from 0.95. Naturally, recalling the earlier success of Newton's laws in uncovering the existence of Neptune and Uranus, the initial blame for the drop was placed on the auxiliary hypotheses. Bayes' theorem can even show us why. Though the numbers in our example are made up, in this case, the auxiliary assumptions were eventually vindicated, and the data about the much greater than expected precession of the perihelion of Mercury undermined Newton's theory, and (as another application of Bayes' theorem would show), increased the probability of Einstein's alternative theory of relativity.

Philosophers and many statisticians hold that the reasoning scientists use to test their hypotheses can be reconstructed as inferences in accordance with Bayes' theorem. These theorists are called Bayesians, and they seek to show that the history of acceptance and rejection of theories in science honors Bayes' theorem, thus showing that in fact, theory testing has been on firm footing all along. Other philosophers and statistical theorists attempt to apply Bayes' theorem in order to determine the probability of scientific hypotheses when the data are hard to get, sometimes unreliable, or only indirectly relevant to the hypothesis under test. For example, they seek to determine the probabilities of various hypotheses about evolutionary events, like the splitting of ancestral species from one another, by applying Bayes' theorem to data about differences in the polynucleotide sequences of the genes of currently living species.

How Much Can Bayes' Theorem Really Help?

How much understanding of the nature of empirical testing does **Bayesianism** really provide? Will it reconcile science's empiricist epistemology with its commitment to unobservable events and processes that explain observable ones? Will it solve Hume's problem of induction? To answer these questions, we must first understand what the probabilities are that all these ps symbolize and where they come from. We need to make sense of p(h), the probability that a certain hypothesis is true. There are at least two questions to be answered: First, there is the "metaphysical" question of what fact it is about the world, if any, that makes a particular probability value, p(h) for a hypothesis, h, the true or correct one.

Second, there is the epistemological question of justifying our estimate of this probability value. The first question may also be understood as a question about the meaning of probability statements, and the second about how they justify inductive conclusions about general theories and future eventualities.

Long before the advent of Bayesianism in the philosophy of science, the meaning of probability statements was already a vexed question. There are some traditional interpretations of probability that we can exclude as unsuitable inter-pretations for the employment of Bayes' theorem. One such is the interpretation of probability as it is supposed to figure in fair games of chance like roulette or blackjack. In a fair game of roulette the chance of the ball landing in any trap is exactly 1/37 or 1/38 because there are 37 (or in Europe 38) traps into which the ball can land. Assuming it is a fair roulette wheel, the probability of the hypoth-esis that the ball will land on number 8 is exactly 1/37 or 1/38 and we know this *a priori*—without experience, because we know *a priori* how many possibilities there are and that each is equally probable (again, assuming the roulette wheel is fair, a bit of knowledge we could never have acquired *a priori* anyway!). But when it comes to hypotheses that can account for a finite body of data, there is no limit to the number of possibilities and no reason to think that each of them has the same probability. Accordingly, the probabilities of a hypothesis about, say, the number of chromosomes in a human nucleus will not be determinable *a priori*, by counting up possibilities and dividing 1 by the number of possibilities.

Another interpretation of probabilities involves empirical observations, for example, coin flips. To establish the frequency with which a coin will come up heads, one flips it several times and divides the number of times it comes up heads by the number of times it was flipped. Will this frequency be a good esti-mate of the probability of heads? It will be when the number of coin flips is large, and the frequencies we calculate for finite numbers of coin flips converge on one value and remain near that value no matter how many times we continue flipping. We can call this value, if there is one, the **long-run relative frequency** of heads. And we treat it as a measure of the probability that the coin comes up heads. But is the long-run relative frequency of heads identical to the probability it will come up heads? This sounds like a silly question, until you ask what the connection is between the long-run relative frequency's being, say 0.5 and the chance that the very next toss will be heads. Notice that a long-run relative fre-quency of 0.5 is compatible with a run of 10, or 100, or 1,000,000 heads in a row, just so long as the total number of tosses is very large, so that a million is a small number in comparison to the total number of tosses. If this is right, the long-run relative frequency is compatible with any finite run of all heads, or all tails, and of course perfectly compatible with the coin's coming up tails on the next toss. Now, suppose we want to know what the probability is that the coin will come up heads on the next toss. If the probability that the coin will come up heads on the next toss is a property of that particular toss, it is a different thing from the long-run relative frequency of heads (which is perfectly compatible with the next 234,382 tosses all being tails). We need some principle that connects the long run to the next toss. One such principle that gets us from the long-run relative

frequency to the probability of the next toss being heads is to assume that coins do in any finite run what they would do in the long run. But this principle is just false. A better principle for connecting long-run relative frequencies to the probability of the next occurrence is something like this: If you know the long-run relative frequency then you know how to bet on whether the coin will land heads or tails, and if you take all bets against heads at odds greater than even money, you will win. But notice this is a conclusion about what you should do as a gambler, not a conclusion about what the coin will in fact do. We will come back to this insight.

Could long-run relative frequencies provide the probability values for a hypothesis without a track record? It is hard to see how. Compare a novel hypothesis to a shiny new penny about to be flipped. Long-run relative frequencies data provide some reason to ascribe a probability of 0.5 to the chances of heads on the new penny. Is there a track record of previous hypotheses relevant to the new one? Only if we can compare it to the right class of similar hypotheses the way we can compare new pennies to old ones. But hypotheses are not like pennies. Unlike pennies, they differ from one another in ways we cannot quantify as we would have to were we to grade them for similarity to one another. Even if we could identify the track record of truth and falsify for similar hypotheses formulated over the past history of science, we would have the problems of (a) justifying the inference from a finite actual sequence to a long-run relative frequency, and (b) justifying the inference from a long-run relative frequency to the next case, the new hypothesis. Recall that in the case of coin flipping, the only connection appears to be that relative frequencies are our best guide to how to lay our bets about the next toss. Perhaps the kind of probability that theory testing invokes is the gambler's kind, what has come to be called "subjective probability." "Subjective" because it reflects facts about the gambler, and what the gambler believes about the past and the future, and "probability" because the bets the gambler makes should honor the axioms of probability.

It is the claim that in scientific testing, the relevant probabilities are subjective probabilities, that is, gambler's odds, which is the distinctive mark of the Bayesian. A Bayesian is someone who holds that at least two of the three probabilities we need to calculate p(h/e) are just a matter of betting odds and that within certain weak constraints, they can take on any values at all. You and I may think that the best betting odds are those that mirror our previous experience of actual frequencies or our estimate of long-run relative frequencies, but this is not part of Bayesianism. The Bayesian holds that in the long run it doesn't matter what values they start with, Bayes' theorem will lead the scientist inexorably to the (available) hypothesis best supported by the evidence. These remarkable claims demand explanation and justification.

Calculating the value of p(e/h) is a matter of giving a number to the probability that e obtains if h is true. This is usually easy to do. If h tells us to expect e, or data close to e, then p(e/h) will be very high. The problem is that using Bayes' theorem also requires that we calculate input values, so called "prior probabilities," p(h) and p(e). p(h) is especially problematical: after all, if h is a new theory

that no one has ever thought of, why should there be any particular right answer to the question of its probability of being true? And assigning a value to p(e), the probability that our data description is correct, may involve so many auxiliary assumptions that even if there is a correct number it is hard to see how we could figure out what it is. The Bayesian asserts that these are not problems. Both values, p(h) and p(e) (and p(e/h) for that matter), are simply degrees of belief, and degrees of belief are simply a matter of what betting odds the scientist would take or decline on whether their beliefs are correct. The higher the odds one takes, the stronger the degree of belief. Here the Bayesian takes a page from economists and others who developed the theory of rational choice under uncertainty. The way to measure a degree of belief is to offer the believer wagers against the truth of his or her belief. Other things being equal, if you are rational, and you are willing to take a bet that h is true at odds of 4:1, then your degree of belief that h is true is 0.8. If you are willing to take odds at 5:1 then your degree of belief is just under 0.9. Probabilities are identical to degrees of belief. The other things that have to be equal for this way of measuring the strength of your beliefs are (a) that you have enough money so that you are not so averse to the risk of losing that it swamps your attraction to the prospect of winning, and (b) that the degrees of belief you assign to your beliefs obey the rules of logic and the three laws of probability above. So long as your degrees of belief, a.k.a. probability assignments, honor these two assumptions, the Bayesian says, the initial values or "prior probabilities" you assign to them can be perfectly arbitrary, but it doesn't really matter. In the parlance of the Bayesians, as more and more data come in, the prior probabilities will be "swamped." That is, when we use Bayes' theorem to "update" prior probabilities, i.e. feed new p(e)s into the latest values for p(e/h) and p(e/h), the successive values of p(h/e) will converge on the correct value, no matter what initial values for these three variables we start with! Prior probabilities are nothing but measures of the individual scientist's purely subjective degree of belief before applying Bayes' theorem. In answer to our metaphysical question about what facts about the world probabilities report, prior probabilities report no facts about the world, or at least none about the world independent of our beliefs. In answer to the epistemological question of what justifies our estimates of probabilities, when it comes to prior probabilities, no more justification is needed or possible than that our estimates obey the axioms of probability.

There is no right or wrong answer as to what the prior probabilities of p(h) or p(e) are, so long as the values of these probabilities obey the rules of probability and logical consistency on betting. Logical consistency simply means that one places one's bets—that is, assigns strengths to one's degrees of belief—in such a way that bookies can't use you for a money pump: someone is a "money pump" for book-makers when they can be made to bet in ways that are irrational (for example, betting that team A will beat team B, team B will beat team C, and team C will beat team A). A series of such bets by one person guarantees that the bookies will make money no matter what team wins. One thing you must do to be sure that you are not a money pump, is apportion your bets in accordance with the axioms of probability theory.

Another theorem of probability theory shows that if we apply Bayes' theorem relentlessly to "update" our prior probabilities as new evidence comes in, the value of p(h) that all scientists assign will converge on a single value no matter where each scientist begins in his or her original assignment of prior probabilities. So not only are prior probabilities arbitrary but it doesn't matter that they are! Some scientists may assign prior probabilities on considerations like simplicity or economy of assumptions, or similarity to already proven hypotheses, or symmetry of the equations expressing the hypothesis. Other scientists will assign prior probabilities on the basis of superstition, aesthetic preference, number worship, or by pulling a ticket out of a hat. It doesn't matter, so long as they all conditionalize on new evidence via Bayes' theorem.

It is not much of an objection to this account of scientific testing that scientists actually offer good reasons for their methods of assigning of prior probabilities. To begin with, Bayesianism doesn't condemn these reasons; at worst it is silent on them. But if features like the simplicity of a hypothesis or the symmetry of its form do in fact increase its prior probability, this will be because a hypothesis having features like this will, via Bayes' theorem, acquire a higher posterior probability than other hypotheses with which it is competing that lack these features. More important, attempts to underwrite the reasoning of scientists who appeal to considerations like economy, simplicity, symmetry, invariance, or other formal features of hypotheses, by claiming that such features increase the objective probability of a hypothesis, come up against the problem that the only kind of probability that seems to make any sense for scientific testing is Bayesian subjective probability. Furthermore, so understood, some Bayesians hold that probabilities can after all deal with some of the traditional problems of confirmation.

One of the major problems confronting Bayesianism, and perhaps other accounts of how evidence confirms theory, is the "problem of old evidence." It is not uncommon in science for a theory to be strongly confirmed by data already well known long before the hypothesis was formulated. Indeed, as we will see in Chapter 14, this is an important feature of situations in which scientific revolutions take place: Newton's theory was strongly confirmed by its ability to explain the data on which Galileo's and Kepler's theories were based. Einstein's general theory of relativity explained previously recognized but highly unexpected data such as the invariance of the speed of light and the precession of the perihelion of Mercury. In these two cases $p(e) = 1$, $p(e/h)$ is very high. Plugging these values into Bayes' theorem gives us

$$p(h/e) = \frac{1 \times p(h)}{1} = p(h)$$

In other words, on Bayes' theorem the old evidence does not raise the posterior probability of the hypothesis—in this case Newton's laws, or the special theory of relativity, at all. Bayesians have gone to great lengths to deal with this problem.

One stratagem is to "bite the bullet" and argue that old evidence does not in fact confirm a new hypothesis. This approach makes common cause with the well-established objection to hypotheses that are designed with an eye to available evidence. Scientists who construct hypotheses by intentional "curve fitting" are rightly criticized and their hypotheses are often denied explanatory power on the grounds that they are *ad hoc*. The trouble with this strategy is that it doesn't so much solve the original Bayesian problem of old evidence as combine it with another problem: how to distinguish cases like the confirmation of Newton's and Einstein's theories by old evidence from cases in which old evidence does not confirm a hypothesis because it was accommodated to the old evidence. The alternative approach to the problem of old evidence is to supplement Bayes' theorem with some rule that gives p(e) a value different from 1. For example, one might try to give p(e) the value it might have had before e was actually observed in the past, or else try to rearrange one's present scientific beliefs by deleting e from them and anything which e makes probable; then go back and assign a value to p(e), which presumably will be lower than 1. This strategy is obviously an extremely difficult one to adopt. And it is (subjectively) improbable that any scientist consciously thinks this way.

Many philosophers and scientists who oppose Bayesianism do so not because of the difficulties that are faced by the program of developing it as an account of the actual character of scientific testing. Their problem is with the approach's commitment to subjectivism. The Bayesian claim that no matter what prior probabilities the scientist subjectively assigns to hypotheses, their subjective probabilities will converge on a single value, is not sufficient consolation to opponents. Just for starters, values of p(h) will not converge unless we start with a complete set of hypotheses that are exhaustive and exclusive competitors. This seems never to be the case in science. Moreover, objectors argue, there is no reason given that the value on which all scientists will converge by Bayesian conditionalization is the *right* value for p(h). This objection of course assumes that there is such a thing as the right, i.e. the objectively correct, probability and so begs the question against the Bayesian. But it does show that Bayesianism is no solution to Hume's problem of induction, as a few philosophers hoped it might be.

And the same pretty much goes for other interpretations of probability. If sequences of events reveal long-run relative frequencies that converge on some probability value and stay near it forever, then we could rely on them at least for betting odds. But to say that long-run relative frequencies will converge on some value is simply to assert that nature is uniform, that the future will be like the past, and so begs Hume's question. Similarly, hypothesizing probabilistic propensities, which operate uniformly across time and space, also begs the question against Hume's argument. In general probabilities are useful only if induction is justified, not vice versa.

There is a more severe problem facing Bayesianism. It is the same problem that we came up against in the discussion of how to reconcile empiricism and explanation in theoretical science. Because empiricism is the doctrine that

knowledge is justified by observation, in general, it must attach the highest probability to statements that describe observations, and lower probability to those that make claims about theoretical entities. Since theories explain observations, we may express the relation between theory and observation as (t and t → h)—where t is the theory and t → h reflects the explanatory relation between the theoretical claims of the theory, t, and an observational generalization, h, describing the data that the theory leads us to expect. The relation between t and h may be logically deductive, or it may be some more complex relation. But p(h) must never be lower than p(t and t → h), just because the antecedent of the latter is a statement about what cannot be observed whose only consequence for observation is h. Bayesian conditionalization on evidence will never lead us to prefer (t and t →h) to h alone. But this is to say that Bayesianism cannot account for why scientists embrace theories at all, instead of just according high subjective probability to the observational generalizations that follow from them.

Of course, if the explanatory power of a theory were a reason for according it a high prior probability, then the embracing of theories by scientists would be rational from the Bayesian point of view. But to accord explanatory power such a role in strengthening the degree of belief requires an account of explanation. And not just any account. It cannot for example make do with the D-N model, for the principal virtue of this account of explanation is that it shows that the explanandum phenomenon could be expected with at least high probability. In other words, it grounds explanatory power on strengthening probability, and so cannot serve as an alternative to probability as a source of confidence in our theories. To argue, as seems tempting, that our theories are explanatory in large part because they go beyond and beneath observations to their underlying mechanisms is something the Bayesian cannot do.

Summary

Empiricism is the epistemology that has tried to make sense of the role of observation in the justification of scientific knowledge. Since the seventeenth century, if not before, a tradition of English-speaking philosophers like Hobbes, Locke, Berkeley, Hume, and Mill have found inspiration in science's successes, and sought philosophical arguments to ground science's claim to special authority in empirical matters. In so doing, these philosophers and their successors set the agenda of the philosophy of science and revealed how complex is the apparently simple and straightforward relation between theory and evidence.

But empiricists have never been uncritical in their assessment of scientific methods and of the epistemological warrant of its claims. We saw some of these problems in connection with the problem of the meaning of theoretical terms and scientific realism in previous chapters. Here we have explored another problem that faces empiricism as the official epistemology of science: the problem of induction, which goes back to Hume, and should be added to the agenda of problems for both empiricists and rationalists.

In the twentieth century the successors of the British empiricists, the logical positivists (or logical empiricists as some of them preferred to be called), sought to combine the empiricist epistemology of their predecessors with advances in logic, probability theory, and statistical inference, to complete the project initiated by Locke, Berkeley, and Hume. In particular, philosophers have appealed to Bayes' theorem, a result provided at the time Hume formulated his problem of induction, to help understand how evidence supports hypotheses in science. But we have seen that appealing to probability is by no means unproblematic. In fact it raises its own problems along with any it might put to rest. We will encounter more of these problems in the next chapter. The problems that empiricism faces in illuminating the epistemology of science continue to mount even as the availability of plausible alternatives declines.

Study Questions

1. Discuss critically: "Lots of scientists pursue science successfully without any regard to epistemology. The idea that science has an 'official one,' and that empiricism is it, is wrong-headed."
2. Why would it be correct to call Locke the father of modern scientific realism and Berkeley the originator of instrumentalism? How would Berkeley respond to the argument for realism as an inference to the best explanation of science's success?
3. What should the relationship be between the ordinary concept of evidence, as for example employed in law courts, and the scientist's use of the term in the testing of general theories?
4. Defend the claim that there are several different but compatible meanings of the word "probability" in science. Is one of them more fundamental than the others?
5. What do you need to add to Bayes' theorem to solve the problem of induction?

Suggested Readings

Gregory Johnson, *Argument and Inference*, is a recent work that expounds inductive logic (as well as deductive reasoning), with exercises for students. A more classical text is Ian Hacking, *Introduction to Probability and Inductive Logic*.

Empiricism is often thought to officially begin with John Locke's *Essay on Human Understanding*. George Berkeley's *Principles of Human Knowledge* is brief but powerful. The last third develops an explicitly instrumental conception of science that he contrasts to Locke's realism. Berkeley argued for idealism—the thesis that only what is perceived exists, that the only things we perceive are ideas, and therefore that only ideas exist. His argument turns on the very same theory

of language that the logical positivists initially embraced: the meaning of every term is given by the sensory idea it names. About Berkeley's work, Hume wrote "it admits no refutation, and carried no conviction" in his *Enquiry Concerning Human Understanding*. In this work he develops the theory of causation discussed in Chapter 2, the theory of language common to empiricists from Berkeley to the logical positivists, and the problem of induction. Bertrand Russell's famous paper, "On Induction," reprinted in Balashov and Rosenberg, brought Hume's argument to center stage in twentieth-century analytical philosophy.

J. S. Mill, *A System of Logic*, carried the empiricist tradition forward in the nineteenth century, and proposed a canon for experimental science still widely employed under the name, Mill's methods of induction. The physicist Ernst Mach, *The Analysis of Sensation*, embraced Berkeley's attack on theory as empirically unfounded against Ludwig Boltzmann's atomic theory. This work was greatly influential on Einstein. In the first half of the twentieth century logical empiricists developed a series of important theories of confirmation, R. Carnap, *The Continuum of Inductive Methods*, H. Reichenbach, *Experience and Prediction*. Their younger colleagues and students wrestled with these theories and their problems.

W. Salmon, *Foundations of Scientific Inference*, is a useful introduction to the history of confirmation theory from Hume through the positivists and their successors. D. C. Stove's *Hume, Probability, and Induction* attempts to solve the problem of induction probabilistically.

L. Savage, *Foundations of Statistics*, provides a rigorous presentation of Bayesianism, as does R. Jeffrey, *The Logic of Decision*. A philosophically sophisticated presentation is P. Horwich, *Probability and Evidence*. An introduction to Bayesianism is to be found in Salmon's *Foundations of Scientific Inference*. Salmon defends the application of the theorem to cases from the history of science in "Bayes' Theorem and the History of Science," reprinted in Balashov and Rosenberg. Richard Swinburne, *Bayes' Theorem*, collects several recent papers on the theorem and its upshot. Important papers on Bayes' theorem and scientific change by Salmon are reprinted in Lange's anthology, and in Curd and Cover's anthology.

The problem of old evidence, among other issues, has led to dissent from Bayesianism by C. Glymour, *Theory and Evidence*. One of his papers on the subject, "Explanation, Tests, Unity and Necessity," is reprinted in Lange, and "Why I Am Not a Bayesian" is reprinted in Curd and Cover.

The philosopher of science Debora Mayo has long argued against the Bayesian approach to hypothesis testing and in favor of the claim that hypotheses need to be subject to "severe tests." See Mayo, *Error and the Growth of Experimental Knowledge*.

P. Achinstein, *The Concept of Evidence*, anthologizes several papers that reflect the complexities of inference from evidence to theory, and the relationship of the notion of evidence to concepts of probability.

B. Skyrms, *From Zeno to Arbitrage: Essays on Quantity, Coherence, and Induction* is a recent advanced treatment of the problems of induction.

11 Confirmation, Falsification, Underdetermination

Overview

Not satisfied with Hume's problem of induction, twentieth-century philosophers have created several more fundamental conceptual problems to be overcome by an empiricist epistemology that grounds the general laws and theories characteristic of so much contemporary science. Among these are Hempel's paradoxes of induction and Goodman's "new riddle of induction." Both show how deeply theoretically entrenched hypothesis testing really is.

At least one important twentieth-century philosopher, Karl Popper, thought he had a way around the problem of induction. Indeed, he thought that the whole problem of building up evidence for a theory represents a deep misunderstanding of what science is all about and how it proceeds. Ironically, the pursuit of his approach to theory testing not only failed to solve the problem but raised so formidable a challenge to empiricism that it produced a movement that simply denies that science is controlled by experience, and even threatens the objectivity of science altogether.

Epistemological Problems of Hypothesis Testing

Assume, along with all working scientists, that either we can solve the problem of induction, or that it is no problem at all (as many philosophers have suggested). Accept that we can acquire knowledge about the future and about laws by experience. Remember this is the claim of the empiricist. It is not a claim about what causes our beliefs about the future and about laws. Everyone will grant that it is experience that does so. No one any longer considers knowledge of how the world works to be innate. Empiricism is a thesis about justification, not (merely) about causation. Experience justifies as well as causes those beliefs that count as knowledge.

A scientific law, even one exclusively about what we can observe, goes beyond the data available, because it makes a claim which, if true, is true everywhere and always, not just in the experience of the scientist who formulates the scientific law. This of course makes science fallible: The scientific law, our current best estimate-hypothesis, may turn out to be (in fact usually does turn out to be) wrong. But it is by experiment that we discover this, and by experiment that we improve on it, presumably getting closer to the natural law that we seek to discover.

It may seem a simple matter to state the logical relationship between the evidence that scientists amass and the hypotheses that evidence tests. But philosophers of science have discovered that testing hypotheses is by no means an easily understood matter. From the outset it was recognized that no general hypothesis of the form "All As are Bs"—for instance, "All samples of copper are electrical conductors"—could be conclusively confirmed because the hypothesis will be about an indefinite number of As and experience can provide evidence about only a finite number of them. By itself a finite number of observations, even a very large number, might be only an infinitesimally small amount of evidence for a hypothesis about a potentially infinite number of, say, samples of copper. At most, empirical evidence supports a hypothesis to some degree. But as we shall see, it may also support many other hypotheses to an equal degree. What is more, and as we have seen is deeply puzzling, often scientists will rightly embrace a hypothesis as expressing a strict law of nature, true everywhere and always, on the basis of a very small number of experiments or observations. The relation of positive evidence to the hypotheses it confirms is obviously complex.

On the other hand, it may seem that such hypotheses could at least be falsified. After all, in order to show that "All As are Bs" is false, one need only find a single A that is not B. One black swan refutes the claim that all swans are white. And understanding the logic of **falsification** is particularly important, because science is fallible. Science progresses by subjecting a hypothesis to increasingly stringent tests, until the hypothesis is falsified, so that it may be corrected, improved, or give way to a better hypothesis. Science's increasing approximation to the truth relies crucially on falsifying tests and scientists' responses to them. Can we argue that while general hypotheses cannot be completely confirmed, they can be completely or "strictly" falsified? Yet it turns out that general hypotheses are not

strictly falsifiable, and this is a fact of the first importance for our understanding of science.

Strict falsifiability is impossible, for nothing follows from a general law alone. From "All swans are white" it does not follow that there are any white swans, because it doesn't follow that there are any swans at all. Recall that Newton's first law may be vacuously true: There are no bodies in the universe free from all forces, owing to the presence of gravitational forces everywhere. To test the generalization about swans we need to independently establish that there is at least one swan and then check its color. The claim that there is a swan, and that we can establish its actual color just by looking at it, are "auxiliary hypotheses" or "auxiliary assumptions." Testing even the simplest hypothesis requires "auxiliary assumptions"—further statements about the conditions under which the hypothesis is tested. For example, to test "All swans are white" we need to establish that "This bird is a swan," and doing so requires we assume the truth of other generalizations about swans besides what their color is. What if the grey bird before us is a grey goose, and not a grey swan? No single falsifying test will tell us whether the fault lies with the hypothesis under test or with the auxiliary assumptions that we need to uncover the falsifying evidence.

To see the problem more clearly consider a test of $PV = nRT$. To subject the ideal gas law to test we measure two of the three variables, say the volume of the gas container and temperature, then use the law to calculate a predicted pressure, and compare the predicted gas pressure to its actual value. If the predicted value is identical to the observed value, the evidence supports the hypothesis. If it does not, then presumably the hypothesis is falsified. But in this test of the ideal gas law we needed to measure the volume of the gas and its temperature. Measuring its temperature requires a thermometer, and employing a thermometer requires us to accept one or more rather complex hypotheses about how thermometers measure heat, for example the scientific law that mercury in an enclosed glass tube expands as it is heated, and does so uniformly. But this is another general hypothesis—an auxiliary we need to invoke in order to put the ideal gas law to the test. If the predicted value of the pressure of the gas diverges from the observed value, the problem may be that our thermometer was defective, or that our hypothesis about how expansion of mercury in an enclosed tube measures temperature change is false. To show that a thermometer was defective, because, say, the glass tube was broken, presupposes another general hypothesis: Thermometers with broken tubes do not measure temperature accurately.

In many cases of testing, the auxiliary hypotheses are among the most basic generalizations of a discipline, like acid turns blue litmus paper red, which no one would seriously challenge. But the logical possibility that they might be mistaken, a possibility that cannot be denied, means that any hypothesis that is tested under the assumption that the auxiliary assumptions are true, can in principle be preserved from falsification by giving up the auxiliary assumptions and attributing any falsity to these auxiliary assumptions. Sometimes, hypotheses are in practice preserved from falsification. Here is a classic example in which the falsification of a test is rightly attributed to the falsity of auxiliary hypotheses

and not the theory under test. In the nineteenth century, predictions of the location in the night sky of Jupiter and Saturn derived from Newtonian mechanics were falsified as telescopic observation improved. But instead of blaming the falsification on Newton's laws of motion, astronomers challenged the auxiliary assumption that there were no other forces, beyond those due to the known planets, acting on Saturn and Jupiter. By calculating how much additional gravitational force was necessary and from what direction, to render Newton's laws consistent with the data apparently falsifying them, astronomers were led to the discovery, successively, of Neptune and Uranus.

As a matter of logic, scientific law can neither be completely established by available evidence, nor conclusively falsified by a finite body of evidence. This does not mean that scientists are not justified on the occasions at which they surrender hypotheses because of countervailing evidence, or accept them because of the outcome of an experiment. What it means is that confirmation and disconfirmation are more complex matters than the mere derivation of positive or negative instances of a hypothesis to be tested. Indeed, the very notion of a positive instance turns out to be a hard one to understand.

Consider the hypothesis that "All swans are white." Suppose we are presented with a white swan and a black boot. Which is a positive instance of our hypothesis? Well, we want to say that only the white bird is; the black boot has nothing to do with our hypothesis. But logically speaking, we have no right to draw this conclusion. For logic tells us that "All As are Bs" if and only if "All non-Bs are non-As." To see this, consider what would be an exception to "All As are Bs." It would be an A that was not a B. But this would also be the only exception to "All non-Bs are non-As." Accordingly, statements of these two forms are logically equivalent. In consequence, all swans are white if and only if all non-white things are non-swans. The two sentences are logically equivalent formulations of the same statement. Since the black boot is a non-white non-swan, it is a positive instance of the hypothesis that all non-white things are non-swans; the black boot is a positive instance of the hypothesis that all swans are white. To many it will seem as though something has gone seriously wrong here. Surely the way to assess a hypothesis about swans is not to examine boots! At a minimum, this result shows that the apparently simple notion of a "positive instance" of a hypothesis is not so simple.

This puzzle is due to Carl G. Hempel and is known as "the paradox of confirmation." There are two broad strategies for dealing with this paradox. Hempel's preferred approach was simply to accept that black boots confirm the hypothesis about all swans being white, and to explain the feeling that they don't do so away as a logically unsophisticated attitude that we can disregard. The other alternative is to argue that if "All swans are white" is a law, it must express some necessary connection between being a swan and whiteness. Recall, this is the explanation of what the law supports, the counterfactual that if my black boot had been a swan it would have been white. If "All swans are white" is an expression of physical or natural necessity, then it will not be logically equivalent to "All non-white things are non-swans," since this statement obviously lacks any

natural or physical necessity. Now a black boot, which is a non-white non-swan, may support this latter general statement, but since it is not equivalent to the law (as it bears no nomic necessity), the black boot won't be a positive instance of the swan hypothesis. This avoids the problem but at the cost of forcing us to take seriously the nature of nomic or physical necessity, which empiricists, and especially logical positivists like Hempel, were reluctant to do.

It is worth noting here that proponents of a Bayesian approach to induction argue that the paradox of confirmation is no problem for Bayesianism. After all, the prior conditional probability of a boot being black, conditional on all swans being white, is lower than the prior probability of the next swan we see being white, conditional on all swans being white. When we plug these two priors into Bayes' theorem, if the prior probabilities of seeing a white swan and a black boot are equal, the probability of "All swans are white" is raised much more by the conditional probability of seeing a white swan, conditional on all swans being white.

But now consider another problem. Consider the general hypothesis that "All emeralds are green." Surely a green emerald is a positive instance of this hypothesis. Now define the term "grue" as "green at time t and t is before 2100 AD or it is blue at t and t is after 2100 AD." Thus, after 2100 AD a cloudless sky will be grue, and any emerald already observed is grue as well. Consider the hypothesis "All emeralds are grue." It will turn out to be the case that every positive instance so far observed in favor of "All emeralds are green" is apparently a positive instance of "All emeralds are grue" as well, even though the two hypotheses are incompatible in their claims about emeralds discovered after 2100 AD. But the conclusion that both hypotheses are equally well confirmed is absurd. The hypothesis "All emeralds are grue" is not just less well confirmed than "All emeralds are green," it is totally without evidential support altogether. But this means that all the green emeralds thus far discovered are not after all "positive instances" of "All emeralds are grue"—else it would be a well-supported hypothesis since there are very many green emeralds and no non-green ones. But if green emeralds are not positive instances of the grue hypothesis, then we need to give a reason why they are not.

One is inclined to respond to this problem by rejecting the predicate "grue" as an artificial, gerrymandered term that names no real property. "Grue" is constructed out of the "real properties" green and blue, and a scientific hypothesis must employ only real properties of things. Therefore, the grue hypothesis is not a real scientific hypothesis and has no positive instances. Unfortunately this argument is subject to a powerful reply. Define bleen as "blue at t and t is earlier than 2100 AD and green at t when t is later than 2100 AD." We may now express the hypothesis that all emeralds are green as "All emeralds are grue at t and t is earlier than 2100 AD or bleen at t and t is later than 2100 AD." Thus, from the point of view of scientific language, "grue" is an intelligible notion. Moreover, consider the definition of "green" as "grue at t and t is earlier than 2100 AD or bleen at t and t is later than 2100 AD." What is it that prevents us from saying that green is the artificial, derived term, gerrymandered from "grue" and "bleen"?

What we seek is a difference between "green" and "grue" that makes "green" admissible in scientific laws and "grue" inadmissible. Following Nelson Goodman, who constructed the problem of "grue," philosophers have coined the term "**projectable**" for those predicates that are admissible in scientific laws. So, what makes "green" projectable? It cannot be that "green" is projectable because "All emeralds are green" is a well-supported law. For our problem is to show why "All emeralds are grue" is not a well-supported law, even though it has the same number of positive instances as "All emeralds are green."

A predicate's being projectable, a general statement's supporting counterfactuals, a regularity's having explanatory power, a predicate-term naming real property, a universal, and a law's being supported by its positive instances—all of these notions turn out to be far more closely tied together than anyone thought.

The puzzle of "grue," known as "the new riddle of induction," remains an unsolved problem in the theory of confirmation. Over the decades since its invention philosophers have offered many solutions to the problem, no one of which has gained ascendancy. But the inquiry has resulted in a far greater understanding of the dimensions of scientific confirmation than the logical positivists or their empiricist predecessors recognized. One thing that all philosophers of science agree on is that the new riddle shows how complicated the notion of confirmation turns out to be, even in the simple cases of generalizations about things we can observe.

Induction as a Pseudo-Problem: Popper's Gambit

Sir Karl Popper was among the most influential of twentieth-century philosophers of science, perhaps even more influential among scientists, especially social scientists, than he was among philosophers. Popper is famous for arguing that Hume's problem of induction is a sort of pseudo-problem, or at least a problem that should not detain either scientists or those who seek to understand the methods of science. Recall that the problem of induction is that positive instances don't seem to increase our confidence in a hypothesis, and the new riddle of induction is that we don't even seem to have a good account of what a positive instance is.

According to Popper, these are not problems for science, since science is not and should not be in the business of piling up positive instances that confirm hypotheses. Popper held that as a matter of fact, scientists seek negative evidence against, not positive evidence for, scientific hypotheses, and that as a matter of method, they are correct to do so. If the problem of induction shows anything, it shows that we should not seek to confirm hypotheses by piling up evidence for them. Instead, good scientific method, and good scientists, seek only to frame substantial conjectures that make strong claims about experience, and then to try as hard as they can to falsify them. After this, scientists should go on to frame new hypotheses and seek their falsification, world without end.

Popper's argument for this methodological prescription (and the descriptive claim that it is what scientists actually do) begins with the observation that in science we seek universal generalizations and that, as a matter of their logical form, "All Fs are Gs" can never be completely confirmed, established, or verified, since the (inductive) evidence is always incomplete. They can, though, as a matter of logic be falsified by only one counterexample. Of course as we have seen, logically speaking, falsification is no easier than verification, owing to the role of auxiliary assumptions required in the test of any general hypothesis. If Popper did not recognize this fact initially, he certainly came to accept that strict falsification is impossible. His claim that scientists do and should seek to frame hypotheses, "conjectures" he called them, and subject them to falsification, "refutation" he sometimes labeled it, must be understood as requiring something different from strict falsification.

Recall in Chapter 3 the example of one sentence expressing more than a single proposition. Depending on its emphasis the sentence "Why did Ms. R. kill Mr. R. with a knife?" can express three distinct questions. Now consider the sentence, "All copper melts at 1,083 degrees centigrade." If we define copper as "the yellowish-greenish metal that conducts electricity and melts at 1,083 degrees centigrade," then of course the hypothesis "All copper melts at 1,083 degrees centigrade" will be unfalsifiable owing to the meanings of the words. Now, suppose you define copper in the same way, except that you strike from the definition the clause about its melting point, and then test the hypothesis. This will presumably eliminate the unfalsifiability due to meaning alone. Now suppose that for many samples you identify as copper, they either melt well below or well above 1,083 degrees centigrade on your thermometer, and in each case you make an excuse for this experimental outcome: the thermometer was defective, or there were impurities in the sample, or it wasn't copper at all, but some similar yellowish-greenish metal, or it was aluminum and illuminated by yellowish-greenish light, or you were suffering from a visual disorder when you read the thermometer, or … The ellipsis is meant to suggest that an indefinitely large number of excuses can be cooked up to preserve a hypothesis from falsification.

Popper argued that such a stratagem—treating a hypothesis as unfalsifiable—is unscientific. Scientific method requires that we envision circumstances which we would count as actually leading us to give up our hypotheses, and that we subject them to test under these conditions. Moreover, Popper argued that the best science is characterized by framing hypotheses that are highly risky—making claims it is easy to test, testing them, and when they fail these tests (as eventually they must), framing new risky hypotheses. Thus, as noted above, he characterized scientific method as "conjectures and refutations" in a book of that title. Like other philosophers of science, including the logical positivists with whom Popper claimed to disagree on most fundamental issues in philosophy, Popper had nothing much to say about the "conjecture" part of science. Philosophers of science have held by and large that there is no logic of discovery, no recipe for how to come up with significant new scientific hypotheses. But Popper did hold that scientists should advance "risky" hypotheses, ones it would

be easy to imagine disconfirming evidence against. And he held that the business of experiment is to seek such disconfirmation.

Popper's claim about falsifiability may be best treated as a description of the attitudes of scientists towards their hypotheses, and/or a prescriptive claim about what the attitudes of good scientists should be, instead of a claim about statements or propositions independent of attitudes towards their testing. It was on this basis that he famously stigmatized Freudian psychodynamic theory and Marx's dialectical materialism as unscientific, employing the possibility of falsification as a criterion to "demarcate" science from pseudoscience. Despite the pretensions of the proponents of these two "theories," neither could be counted as scientific, for as "true believers" their proponents would never countenance counterexamples to them that require the formulation of new conjectures. Therefore, Popper held that their beliefs were not properly to be considered scientific theories at all, not even repudiated ones. At one point Popper also treated Darwin's theory of natural selection as unfalsifiable, owing in part to the proclivity of biologists to define fitness in terms of reproductive rates and so turn the PNS (see Chapter 9) into a definition. Even when evolutionary theorists are careful not to make this mistake, Popper held that the predictive content of adaptational hypotheses was so weak that falsification of the theory was impossible.

Since repudiating Darwin's theory was hardly plausible, Popper allowed that though it was not a scientific theory strictly speaking, it was a valuable metaphysical research program. Of course Marxian and Freudian theorists would have been able to make the same claim. More regrettably, religiously inspired opponents of the theory of natural selection were only too happy to cloak themselves in the mantle of Popper: they argued that either Christian metaphysics had to share equal time with Darwinian metaphysics in science classrooms, or the latter should be banished along with the former. It is worth noting for the record that Darwin faced the challenge that Popper advances, of identifying circumstances that would falsify his theory, in Chapter 6 of *On the Origin of Species*, entitled "Difficulties of the Theory."

Stigmatizing some theories as pseudoscience was subsequently adopted, especially by economic theorists. This may well have been because of Popper's personal influence on them, or owing to his other writings that attacked Marxist political economy and political philosophy. Many social scientists made common cause with Popper on this account. The embrace of Popper by economic theorists particularly was ironic in two respects. First, their own practice completely belied Popper's maxims. For more than a century economic theorists (including the Popperians among them) had been utterly committed to the generalization that economic agents are rational-preference maximizers, no matter how much evidence behavioral, cognitive and social psychologists built up to disconfirm this generalization. Second, in the last two decades of the twentieth century the persistence in this commitment to the economic rationality of consumers and producers, despite substantial counterevidence, eventually paid off. The development of game theory, and especially evolutionary game theory,

vindicated the economists' refusal to give up the assumption of rationality in spite of alleged falsifications.

What this history shows is that at least when it comes to economics, Popper's claims seem to have been falsified as descriptions and to have been ill advised as prescriptions. The history of Newtonian mechanics offers the same verdict on Popper's prescriptions. It is a history in which for long periods scientists were able to reduce narrower theories to broader theories, while improving the predictive precision of the narrower theories, or showing exactly where these narrower theories went wrong, and were only approximately correct. The history of Newtonian mechanics is also the history of data forcing us to choose between "ad hoc" adjustments to auxiliary hypotheses about initial conditions and falsifying Newtonian mechanics, in which apparently the "right" choice was preserving the theory. Of course sometimes, indeed often, the right choice is to reject theory as falsified, and frame a new hypothesis. The trouble is to decide in which situation scientists find themselves. Popper's one-size-fits-all recipe, "refute the current theory and conjecture new hypotheses," does not always provide the right answer.

The history of physics also seems to provide counterexamples to Popper's claim that science never seeks, nor should it seek, confirmatory evidence for a theory. In particular scientists are impressed with "novel" predictions, cases in which a theory is employed to predict a hitherto completely undetected process or phenomenon, and even sometimes to predict its quantitative dimensions. Such experiments are treated not merely as attempts to falsify that fail, but as tests that positively confirm.

Recall the problems that physicists and empiricists had with Newton's occult force of gravity. In the early twentieth century Albert Einstein advanced a "general theory of relativity" that provided an account of motion that dispensed with this. Einstein theorized that there is no such thing as gravity (some of his arguments were methodological, or philosophical). Instead, Einstein's theory holds, space is "curved," and more steeply curved around massive bodies like stars. One consequence of this theory is that the path of photons should be bent in the vicinity of such massive bodies. This is not something that Newton's theory should lead us to expect, since photons have no mass and so are not affected by gravity—recall the inverse square law of gravitational attraction in which the masses of bodies gravitationally attracting one another affect the force of gravity between them. In 1919, at great expense, a British expedition was sent to a location in South America where a total solar eclipse was expected, in order to test Einstein's theory. By comparing the apparent location in the sky of stars the night before the eclipse and their apparent location during the eclipse (when stars are visible as a result of the Moon's blocking the Sun's normal brightness in the same region of the sky), the British team reported the confirmation of Einstein's hypothesis. The result of this test and others was of course to replace Newton's theory with Einstein's.

Many scientists treated the outcome of this expedition's experiment as strong confirmation of the general theory of relativity. While applauding the riskiness

of Einstein's prediction, thus opening it to the possibility of falsification, Popper would of course have to insist that they were mistaken. At most, the test falsified Newton's theory, while leaving Einstein's unconfirmed. One reason that many scientists would reject this claim is that in the subsequent 80 years, as new and more accurate devices became available for measuring this and other predictions of Einstein's theory, its consequences for well-known phenomena were confirmed to more and more decimal places, and more important, its novel predictions about phenomena no one had ever noticed or even thought of were confirmed. Still, Popper could argue that scientists were mistaken in holding the theory to be confirmed. After all, even if the theory does make more accurate predictions than Newton's, they don't match up 100 percent with the data, and excusing this discrepancy by blaming the difference on observational error or imperfections in the instruments is just an *ad hoc* way of preserving the theory from falsification. One thing Popper could not argue is that the past fallibility of physics shows that probably Einstein's general theory of relativity is also at best an approximation and not completely true. Popper could not argue this way, for this is an inductive argument, and Popper agrees with Hume that such arguments are ungrounded.

What can Popper say about theories that are repeatedly tested, whose predictions are borne out to more and more decimal places, which make novel, striking predictions that are in agreement with (we can't say "confirmed by") new data? Popper responded to this question by invoking a new concept: "corroboration." Theories can never be confirmed, but they can be corroborated by evidence. How does corroboration differ from confirmation? It is a quantitative property of hypotheses, which measures their content and testability, their simplicity, and their previous track record of success in standing up to attempts to falsify them in experiments. For present purposes, the details of how corroboration differs from confirmation is not important, except that corroboration cannot be a relationship between a theory and already available data that either (a) makes any prediction about future tests of the theory, or (b) gives us any positive reason at all to believe that the theory is true or even closer to the truth than other theories. The reason is obvious. If corroboration had either of these properties it would be at least in part a solution to the problem of induction, and this is something Popper began by dispensing with.

If hypotheses and theories are the sorts of things that people can believe to be true, then it must make sense to afford them more credibility than others, as more reasonable to believe than others. It may well be that among the indefinitely many possible hypotheses, including all the ones that never have and never will occur to anyone, the theories we actually entertain are less well supported than others, are not even approximately true, and are not improving in approximate truth over their predecessors. This possibility may be a reason to reject increasing confirmation as merely shortsighted speculation. But it is an attitude difficult for working scientists to take seriously. As between competing hypotheses they are actually acquainted with, the notion that none is more reasonable to believe than any other doesn't seem attractive. Of course, an instrumentalist

about theories would not have this problem. On the instrumentalist view, theories are not to be believed or disbelieved, they are to be used when convenient, and otherwise not. Instrumentalists may help themselves to Popper's rejection of induction in favor of falsification. But, ironically, Popper was a realist about scientific theories.

Underdetermination

The testing of claims about unobservable things, states, events, and processes is evidently a complicated affair. In fact the more one considers how observations confirm hypotheses, and how complicated the matter is, the more one is struck by a certain inevitable and quite disturbing "**underdetermination**" of theory by observation.

As we have noted repeatedly, the "official epistemology" of modern science is empiricism—the doctrine that our knowledge is justified by experience—observation, data collection, and experiment. The objectivity of science is held to rest on the role that experience plays in choosing between hypotheses. But if the simplest hypothesis comes face to face with experience only in combination with other hypotheses, then a negative test may be the fault of one of the accompanying assumptions, a positive test may reflect compensating mistakes in two or more of the hypotheses involved in the test that cancel one another out. Moreover, if two or more hypotheses are always required in any scientific test, then when a test prediction is falsified there will always be two or more ways to "correct" the hypotheses under test. When the hypothesis under test is not a single statement like "All swans are white" but a system of highly theoretical claims like the kinetic theory of gases, it is open to the theorist to make one or more of a large number of changes in the theory in light of a falsifying test, any one of which will reconcile the theory with the data. But the large number of changes possible introduces a degree of arbitrariness that seems foreign to our picture of science. Start with a hypothesis constituting a theory that describes the behavior of unobservable entities and their properties. Such a hypothesis can be reconciled with falsifying experience by making changes in it that cannot themselves be tested except through the same process all over again—one which allows for a large number of further changes in case of falsification. It thus becomes impossible to establish the correctness, or even the reasonableness, of one change over another. Two scientists beginning with the same theory, subjecting it to the same initial disconfirming test, and repeatedly "improving" their theories in the light of the same set of further tests will almost certainly end up with completely different theories, both equally consistent with the data their tests have generated. If it is empirical data and only empirical data that test theory, and if empirical data do not point to where theories that are disconfirmed need to be changed, then over time, theories in the sciences should continually proliferate. But this does not appear to be the case, especially in the physical sciences. We return to this point and its implications below.

The problem of empirically equivalent but logically incompatible theories becomes especially serious as science becomes more theoretical. A famous example of van Fraassen's illustrates this point. Recall van Fraassen's view, "constructive empiricism," which urges that we be agnostic about the truth claims of the theoretical "parts" of theories. One of his arguments rests on the following possibility: Compare Newtonian mechanics—the four laws we discussed in Chapter 7 that explain so much in physics. Add to these four laws the further axiom that the universe and everything in it is moving at 100 kilometers per hour along a vector whose direction is given by extending a line from the Earth to the North Star, Polaris. These two theories will be equally powerful in explanation and prediction; there will be no empirical way to tell the difference between them. They are empirically equivalent. Which means that our choice between them is underdetermined by observations.

Suppose one objects to this example on the grounds—well established by Einstein's special theory and before that argued by several philosophers from Leibniz and Berkeley in the eighteenth century onwards—that there is no such thing as motion in absolute space. Since there is empirical evidence that falsifies this assumption, we cannot add to Newton's laws to produce an empirically equivalent theory. This line of argument has several problems. First, if it relies on the strict falsification of the assumption of motion in a direction in absolute space, then it begs the question against underdetermination. Second, it is only retrospectively that we know the additional assumption to be factually false. Since we cannot know which are the false parts of theories that are empirically equivalent with respect to evidence available at a given time, as opposed to later, we cannot use this objection to deny the possibility of underdetermination before the fact. Third, the example van Fraassen constructed is meant to illustrate a possibility that for all we know could have obtained in the past and may well obtain in the future.

Imagine, now, the "end of inquiry" when all the data on every subject are in. Can there still be two distinct, equally simple, elegant, and otherwise satisfying theories that are equally compatible with all the data, and incompatible with one another? Given the empirical slack present even when all the evidence appears to be in, the answer seems to be that such a possibility cannot be ruled out. Since they are distinct theories about everything, our two total "systems of the world" must disagree somewhere, they must be incompatible, and therefore cannot both be true. We cannot either remain agnostic about whether one is right or ecumenical about embracing both. Yet it appears that observation would not be enough to decide between these theories.

Contemporary debates in cosmology may illustrate the possibility we contemplate here. Worse, they may exemplify its actuality. There are several versions of a superstring "theory of everything" that combine both quantum mechanics and the general theory of relativity, and between which it may be impossible to choose empirically, since the observations needed to do so require at least as much energy as there is in the universe. In addition there are non-string theory alternatives, so-called quantum-loop gravity theories, that are also at present

empirically equivalent, and may for all we know be permanently empirically equivalent.

If it is possible for there to be, at the actual or merely hypothetical "end of inquiry," more than one empirically adequate theory, equally compatible with all the evidence, equally general in explanatory domain, and equally precise in predictive power, then clearly empiricism is in serious trouble. As total theories, they cannot both be true for they will disagree about something theoretical. Yet as empirically equivalent, no data will be able to choose between them. If there is a fact of the matter about which one is true, it will not be accessible on any empiricist epistemology.

Even if we find a way to rule out incompatible, empirically equivalent, equally powerful total theories, we still face a severe problem of the underdetermination by observation of the actual theories that scientists have adopted. And yet science does not show the sort of proliferation of theories that this empirical underdetermination would lead one to expect. The kind of intractable, irresolvable theoretical disputes that underdetermination makes possible are almost never actual. One might argue that this fact reflects a lack of imagination, a failure of scientists to conceive, contemplate or explore alternative empirically adequate theories. (Indeed one might count this as a point for antirealists to make against claims that without realism the success of science would be a miracle (explored in Chapter 8). Rather than presume that the success of our theories vindicates science's convergence on truth, perhaps it merely demonstrates a failure of imagination by scientists to explore enough empirically equivalent theories.)

The more we consider reasons why this sort of underdetermination does not manifest itself, the more problematical becomes the notion that scientific theory is justified by objective methods that make experience the final court of appeal in the certification of knowledge. For what else besides the test of observation and experiment could account for the theoretical consensus that is characteristic of most natural sciences? Of course there are disagreements among theorists, sometimes very great ones, yet over time these disagreements are settled, to almost universal satisfaction. If, owing to the ever-present possibility of underdetermination, this theoretical consensus is not achieved through the "official" methods, how is it achieved?

Besides the test of observation, theories are also judged on other criteria: simplicity, economy, explanatory unification, precision in prediction, and consistency with other already adopted theories. In theory choice we are not limited merely to the derivation of predictions for observations that we can test. When observations disconfirm a body of hypotheses, there are methodological guidelines that enable us to design new experiments and tests, which may enable us to point the finger more precisely at one or another of the components of our original package of hypotheses. And here too considerations of simplicity, explanatory unification, precision of prediction, amount of allowable experimental error, and consistency with other established theories again apply. Theory

choice is a continual process of iterative application of this same toolbox of considerations in order to assess the implications of empirical observation in making theory choices.

Theory choice sounds like it is controlled by empirical observation after all, though ones that are guided by broad rules and regulations of empirical inquiry. But what are the grounds of these rules and regulations? Two obvious answers suggest themselves, neither satisfactory. First, the methodological principles we add to observation in order to eliminate the threat of underdetermination could be vouchsafed to us by some sort of *a priori* considerations. Rationalists like Kant thought that there were such criteria for theory choice that guaranteed the truth of Newton's theory against skepticism. Clearly, this source of assurance against the threat of underdetermination won't be tolerated by empiricism. But the alternative looks question-begging.

Suppose empiricists argue that the extra-observational criteria of theory choice have themselves been vindicated by observation and experiment. But then these criteria would be guilty of simply invoking observations—albeit somewhat indirectly. A theory's consistency with other already well-established theories confirms that theory only because observations have established the theories it is judged to be consistent with. Simplicity and economy in theories are themselves properties that we have observed nature to reflect and other well-confirmed theories to bear, and we are prepared to surrender them if and when they come into conflict with our observations and experiments. An empiricist justification for the methodological rules we employ in theory choice is circular as an argument against the threat of underdetermination.

Having excluded both rationalist and empiricist sources for the consensus that science has shown in the face of underdetermination, over the last 400 years especially, philosophers of science face a serious problem.

There is one alternative source that almost all philosophers of science are strongly disinclined to accept: the notion that theoretical developments are epistemically guided by non-experimental, non-observational considerations, such as *a priori* philosophical commitments, religious doctrines, political ideologies, aesthetic tastes, psychological dispositions, social forces, or intellectual fashions. Such factors we know will make for consensus, but not necessarily one that reflects increasing approximation to the truth, or to objective knowledge. Indeed, these non-epistemic, non-scientific forces and factors are supposed to deform understanding and lead away from truth and knowledge.

The fact remains that a steady commitment to empiricism, coupled with a fair degree of consensus about the indispensability of scientific theorizing, strongly suggests the possibility of a great deal of slack between theory and observation. But the apparent absence of arbitrariness fostered by underdetermination demands explanation. And if we are to retain our commitment to science's status as knowledge *par excellence*, this explanation had better be one we can parlay into a justification of science's objectivity as well. The next chapter shows that prospects for such an outcome are clouded with doubt.

Summary

In this chapter we started out with what looked like small problems, cute paradoxes, clever philosophical conundrums about white swans and black boots and funny properties that seem to involve changing colors, but these matters ended up making very serious trouble for empiricism.

Attempts like Popper's to entirely short-circuit inquiry about how evidence supports theory and to replace it with the question of how it falsifies theory, make matters worse. For they seem to boomerang into the persistent possibility of underdetermination, and the global threat that it may not be observation, experiment, and data collection that really control inquiry. In other words, a series of problems confronting empiricism has become increasingly formidable. This is especially so when we consider that these problems for empiricism have in fact surfaced, and been taken seriously, largely by empiricists seeking to ground ever more firmly their epistemology.

As we have seen, given the role of auxiliary hypotheses in any test of a theory, it follows that no single scientific claim meets experience for test by itself. It does so only in the company of other, perhaps large numbers of other hypotheses needed to effect the derivation of some observational prediction to be checked against experience. But this means that a disconfirmation test, in which expectations are not fulfilled, cannot point the finger of falsity at one of these hypotheses and that adjustments in more than one may be equivalent in reconciling the whole package of hypotheses to observation.

As the size of a theory grows, and it encompasses more and more disparate phenomena, the alternative adjustments possible to preserve or improve it in the face of recalcitrant data increase. Might it be possible, at the never actually to be reached "end of inquiry" (when all the data are in), that there be two distinct total theories of the world, both equal in evidential support, simplicity, economy, symmetry, elegance, mathematical expression, or any other desideratum of theory choice? A positive answer to the question may provide powerful support for an instrumentalist account of theories. For apparently there will be no fact of the matter accessible to inquiry that can choose between the two theories.

Yet, the odd thing is that underdetermination is a mere possibility. In point of fact, it almost never occurs. This suggests two alternatives. The first alternative, embraced by most philosophers of science, is that observation really does govern theory choice (else there would be more competition among theories and models than there is); it's just that we simply haven't figured it all out yet. The second alternative is more radical, and is favored by a generation of historians, sociologists of science, and a few philosophers who reject both the detailed teachings of logical empiricism and also its ambitions to underwrite the objectivity of science. On this alternative, observations underdetermine theory, but it is fixed by other facts—non-epistemic ones, like bias, faith, prejudice, and the desire for fame or at least security, power politics. This radical view, that science is a process, like other social processes, and not a matter of objective progress, is the subject of the next two chapters.

Study Questions

1. Defend or criticize: "We all know a positive instance when we see one. No one need worry about the paradox of confirmation. The same goes for telling projectable predicates like 'green' from unprojectable ones like 'grue.'"
2. What's wrong with this claim: "It's always clear in science which statement is falsified by a disconfirmed prediction, and it's never the claims about the apparatus. 'A poor carpenter always blames his tools' is a sound maxim in experimental science."
3. Why can't we always claim that two equally well-confirmed total theories that appear to be incompatible are only disguised terminological variants of one another?
4. Exactly why is underdetermination a real threat to empiricism?
5. Is underdetermination a real threat to the objectivity of science?

Suggested Readings

The paradoxes of confirmation first broached by Hempel in his essays on confirmation theory collected in *Aspects of Scientific Explanation* are of special importance, as is N. Goodman, *Fact, Fiction and Forecast*, where the new riddle of induction is introduced along with Goodman's path-breaking treatment of counterfactuals. Hempel's and Goodman's papers are anthologized in Lange. Peter Achinstein's paper, "The Grue Paradox," which appears in print initially in Balashov and Rosenberg, is an invaluable exposition of Goodman's new riddle, and a novel solution.

Popper, "Science: Conjectures and Refutations," from the book of the same name is reprinted in Curd and Cover, as his attack on "The Problem of Induction."

The possibility of underdetermination was recognized early in the twentieth century by the French philosopher Pierre Duhem, in *The Aim and Structure of Physical Theory*, and later made central to the work of Quine, *Word and Object*. His work is discussed at length in Chapter 14. The problem of underdetermination was subject to sustained critical scrutiny over the succeeding half-century. A paper by Duhem is reprinted in Curd and Cover. For an important example of this criticism, see J. Leplin and L. Laudan, "Empirical Equivalence and Underdetermination," and C. Hoefer and A. Rosenberg, "Empirical Equivalence, Underdetermination and Systems of the World," who respond to their denial of underdetermination. Laudan, "Demystifying Underdetermination," is reprinted in Curd and Cover.

An excellent recent treatment of problems of underdetermination is Stanford, *Exceeding Our Grasp*.

12 Challenges from the History of Science

Overview

If observational evidence underdetermines theories, we need an explanation at least of what does determine the succession of theories that characterize science's history. Even more, for philosophy's purposes we need justification for the claim that these observationally unsupported theories are epistemically rational and reasonable ones to adopt. Clearly, empiricism cannot do this by itself, as its resources in justification are limited to observation.

Thomas Kuhn, an important historian and philosopher of science, was among the first to search the history of science for these non-observational factors that explain theory choice, and to consider how they might justify it as well. His book, *The Structure of Scientific Revolutions*, sought to explore the character of scientific change—how theories succeed one another— with a view to considering what explains and justifies the replacement of one theory by another. The logical positivists held that theories are chosen on the basis of observation and succeed one another by reduction, which preserves what is correct in an earlier theory and so illuminates the history of science as progress. Kuhn's research challenges both of these ideas.

By introducing considerations from psychology and sociology as well as history, Kuhn reshaped the landscape in the philosophy of science and made it take seriously the idea that science is not a disinterested pursuit of the truth, successively cumulating in the direction of greater approximation to it, as guided by unambiguous observational test.

Kuhn's shocking conclusions suggest that science is as creative an undertaking as painting or music, which encouraged many to view science as no more objectively progressive, correct, or approximately true about the world than other human activities. On this view, the history of science is the history of change, but not progress. We are no nearer the truth about the nature of things nowadays than we were in Aristotle's time. These surprising conclusions represent a great challenge to contemporary philosophy of science.

Some philosophers have responded to Kuhn's work by seeking explicitly to show that its history is one of rational progress. This chapter concludes by exploring the approach of one of the most influential of these philosophers, Imre Lakatos.

A Role for History in the Philosophy of Science?

In the last several chapters we traced the problems confronting philosophy's traditional analysis of scientific knowledge as the outcome of attempts to explain our observations that are themselves "controlled" by our observations. Empiricism, the ruling "ideology" of science, assures us that what makes scientific explanations credible, and what ensures the self-correction of science as well as its ever-increasing predictive powers, is the role that observation, experiment, and test play in the certification of scientific theory.

We have seen that making this role precise is not something the philosophy of science has been able to do. Not only can the philosophy of science not provide an uncontroversial empiricist justification for our knowledge of the existence of theoretical entities, it cannot even assure us that the terms that name these entities are meaningful. Even worse, the simplest evidential relation between a piece of data and a hypothesis which that data might test seems equally difficult to express with the sort of precision that both science and the philosophy of science seem to require. One might hold that this is not a problem for scientists, just for philosophers of science. After all, we know that theoretical terms are indispensable because theoretical entities exist and we need to invoke them in explanations and predictions. And we know that the abilities of scientific hypotheses to withstand empirical test is what makes them knowledge. Formalizing these facts may be an interesting exercise for philosophy but it need not detain the working scientist.

This would be a superficial view of the matter. To begin with, it would be a double standard not to demand the same level of detail and precision in our understanding of science that science demands of itself in its understanding the world. Scientific empiricism bids us to test our ideas against experience; we cannot do this if these ideas are vague and imprecise. The same must go for our ideas about the nature of science itself. Second, if we cannot provide a precise and detailed account of such obvious and straightforward matters as the existence of theoretical entities and the nature of scientific testing, then this is a

symptom that there may be something profoundly wrong in our understanding of science. This will be of particular importance to the extent that less well-developed disciplines look to the philosophy of science for guidance, if not recipes on how to be scientific.

The dissatisfaction with philosophy of science's answers to fundamental questions about theories and their testing of course led philosophers of science to begin rethinking the most fundamental presuppositions of logical empiricism. This re-examination began with the uncontroversial claim that the philosophy of science should provide a picture of the nature of science that mirrors what we know about its history and its actual character. This may sound uncontroversial until it is recalled how much traditional philosophy of science relied on considerations from formal logic coupled with a narrow range of examples from physics.

One of the earliest and certainly the most influential document in the reconsideration of the nature of science from the perspective of its history was Thomas Kuhn's *Structure of Scientific Revolutions*. This slim work set out to bring the philosophy of science face to face with important episodes from its history. But it ended up completely undermining philosophy's confidence that it understood anything about science. And it became the single most heavily cited work in the second half of the twentieth century's absorption with science. How could this have happened?

Kuhn's study of the history of science from well before Newton suggested to him that claims about the world (that we might now view as pre- or unscientific myths) were embraced by learned people whose aim was to understand the world for much the same sorts of reasons that we embrace contemporary physical theory. If it is these sorts of reasons that support a belief that make it scientific, then these myths were science, too. Alternatively, our latest scientific beliefs could be myths, like the pre- and unscientific ones they replaced. Kuhn held that the first of these alternatives was to be preferred. Adopting this perspective makes the history of long-past science an important source of data in any attempt to uncover the methods that make science objective knowledge. The second alternative, that contemporary science is just the latest successor in a sequence of mythic "world-views," no more "objectively true" than its predecessors, seemed to most philosophers of science (if not always to Kuhn) preposterous. The trouble is that Kuhn's account of the nature of science was widely treated outside philosophy of science as having supported this second alternative at least as much as the first.

Kuhn's ostensible topic was scientific change, how the broadest theories replace one another during periods of scientific revolution. Among the most important of these was the shift from Aristotelian physics to Newtonian mechanics, from phlogiston chemistry to Lavoisier's theories of reduction and oxidation, from non-evolutionary biology to Darwinism, and from Newtonian mechanics to relativistic and quantum mechanics. Periods of revolutionary change in science alternate with periods of what Kuhn called "**normal science**," during which the direction, methods, instruments, and problems that scientists face are all fixed by the established theory. But Kuhn thought that the term "theory" did not

aptly describe the intellectual core of a program of "normal science." Instead he coined the term "**paradigm**," a word which has gone into common usage so broadly that few appreciate its origin in a work of academic history and philosophy of science. To employ terms from subsequent paradigms that have taken over culture, "paradigm" is a meme that has gone viral.

Paradigms are more than just equations, laws, or statements encapsulated in the chapters of a textbook. The paradigm of Newtonian mechanics was not just Newton's laws of motion, it was also the model or picture of the universe as deterministic clockwork in which the fundamental properties of things were their position and momentum, from which all the rest of their behavior could eventually be derived. The Newtonian paradigm also included the standard set of apparatus or lab equipment whose behavior was explained, predicted and certified by Newton's laws, and with it a certain strategy of problem-solving. The Newtonian paradigm includes a methodology, a philosophy of science, indeed an entire metaphysics. In his later writing Kuhn placed more emphasis on the role of the **exemplar**—the apparatus, the practice, the impedimenta—of the paradigm than on any verbal expression of its content. The exemplar more than anything defines the paradigm.

Paradigms drive normal science, and this kind of science is in a crucial way quite different from the account of it advanced by empiricist philosophers of science. Instead of following where data, observation, and experiment lead, Kuhn held that normal science dictates the direction of scientific progress by determining what counts as an experiment and when observations need to be corrected to count as data. During normal science, research focuses on pushing back the frontiers of knowledge by applying the paradigm to the explanation and prediction of data. What it cannot explain is anything outside of its intended domain. Within its domain what it cannot predict is plain old experimental error or the clumsy misapplication of the paradigm's rules by a scientist who has not fully understood the paradigm.

Under the auspices of normal science, three sorts of empirical enquiries flourish: those which involve redetermining previously established observational claims to greater degrees of precision; the establishment of facts without significance or importance for themselves but which vindicate the paradigm; experiments undertaken to solve problems to which the paradigm draws our attention. Failure to accomplish any of these three aims reflects on the scientist attempting them, not the paradigm employed. None of these sorts of inquiry is to be understood on the empiricist model of experience as testing theory.

The grandest example of the success of normal science in giving priority to theory over data (and thus undermining empiricism) is found in the story of Newtonian mechanics and the planets Neptune and Uranus. One of the great successes of Newtonian mechanics in the 1700s was predicting the appearance and reappearance of Halley's comet by enabling astronomers to calculate its orbit. In the nineteenth century, apparent improvements in telescopes enabled astronomers to collect data on the path of Saturn that suggested an orbit different from that which Newtonian theory predicted. But this apparently

falsifying observation discredits a "package" of hypotheses, one that includes Newton's laws, plus a large number of auxiliary hypotheses about how telescopes work, and what corrections have to be made to derive data from observations using them, as well as assumptions about the number and mass of the known planets whose forces act upon Saturn. The falsifying observations seem at first blush to underdetermine which component of the package should be surrendered. But the centrality of the Newtonian paradigm to normal science in physics did not in fact leave matters underdetermined. Instead the ruling paradigm dictated that the data on Saturn be treated as a "puzzle," that is, a problem with a "correct" answer to be discovered by the ingenuity of physicists and astronomers. A physicist's failure to solve a problem within the paradigm simply discredited the physicist, not the physicist's paradigm! There could be no question that the theory was wrong; it had to be the instruments, the astronomers, or the assumptions about the number and mass of the planets. And indeed, this was how matters turned out. Accepting the force of the Newtonian paradigm, and the reliability of the instruments that the Newtonian paradigm certified, left only the option of postulating one or more additional planets (as yet undetected because too small or too distant or both), whose Newtonian gravitational forces would cause Saturn to move in the way that the new data suggested. Training their telescopes in the direction from which such forces must be exerted, astronomers eventually discovered first Neptune and then Uranus, thus solving the puzzle set by the Newtonian paradigm. Whereas the empiricist would describe the outcome as an important empirical confirmation of Newton's theory, followers of Kuhn would insist that the paradigm was never in doubt and so neither needed nor secured additional empirical support from the solution to the puzzle.

Normal science is characterized by textbooks, which despite their different authors convey largely the same material, with the same demonstrations, experiments, and similar lab manuals. Normal science's textbooks usually contain the same sorts of problems at the back of each chapter. Solving these puzzles in effect teaches scientists how to treat their subsequent research agenda as sets of puzzles. Naturally, some disciplines are, as Kuhn put it, "pre-paradigmatic," as evinced for example by the lack of textbook uniformity. These disciplines are ones, like many of the social sciences (but not economics), where the lack of commonality among the textbooks reveals an absence of consensus. How the competition in pre-paradigm science gives way to a single winner, which then determines the development of normal science, Kuhn does not tell us. But he does insist that paradigms do not triumph by anything like what the experimental method of empiricism suggests. And the reason Kuhn advances is an epistemologically radical claim about the nature of observation in science.

Recall the distinction between observational terms and theoretical terms so important to the project of empiricism. According to empiricism, observational terms are used to describe the data that epistemically control a theory. The empiricist's problem is that observation seems inadequate to justify the

explanatory theories about unobservable events, objects, and processes with which science explains the observable regularities we experience in the lab and world. This problem for empiricism is not a problem for Kuhn, because he denies that there is a vocabulary that describes observations that is neutral between competing theories. According to Kuhn, paradigms extend their influence not just to theory, philosophy, methodology, and instrumentation, but also to the lab-bench and the field notebook, dictating observations, not passively receiving them.

Kuhn cited evidence from psychological experiments about optical illusions, gestalt-switches, expectation-effects, and the unnoticed theoretical commitments of many apparently observational words we incautiously suppose to be untainted by presuppositions about the world. Consider some examples. Kuhn's example was a red jack of spades and a black jack of hearts that most people don't notice as they are accustomed to black spades and red hearts. Since Kuhn first made the point, other examples have become common knowledge. In the Müller-Lyer illusion, two lines of equal length, one with arrows at each end pointing out, and the other with arrows pointing in, are viewed by Western eyes as unequal, but the illusion does not fool people from "non-carpentered societies" without experience of straight lines. The Necker cube, a simple two-dimensional rendering of a transparent cube, is not so identified by those without experience of perspective, and the front-back switch or reversal which we can effect in our perception shows that the act of seeing is not a cognitively innocent one. When Galileo first described the moon as "cratered," his observations already presupposed a minimal theoretical explanation of how the lunar landscape was created—by impacts from other bodies.

Kuhn was not alone in coming to this conclusion. Several opponents of empiricism in the 1950s came to hold this view about observation. They held that the terms in which we describe observations, whether given by ordinary language or scientific neologisms, presuppose divisions or categorizations of the world of experience in ways that reflect prior "theories": the categories we employ to classify things, even categories as apparently theory-free as color, shape, texture, sound, taste, not to mention size, hardness, warmth/coldness, conductivity, transparency, etc., are shot through with interpretation. Observation, that is, is theory-laden. Instead of seeing a glass of milk, we see "it" as a glass of milk, where the "it" is not something we can describe separately in a theory-neutral vocabulary. Even the words "white," "liquid," "glass," "wet," "cold," or however we seek to describe our sensory data, are as much theory-bound as "magnetic" or "electric" or "radioactive."

Since Kuhn first wrote, this claim that the theoretical/observational distinction is at least unclear and perhaps baseless has become a lynch-pin for non-empiricist philosophy of science. Its impact upon the debate about the nature, extent, and justification of scientific knowledge cannot be overstated. In particular it makes it much more difficult to understand the nature of scientific testing—the most distinctive of science's differences from everything else. Kuhn recognized this consequence, and his way of dealing with it is what made *The Structure of Scientific Revolutions* so influential.

New Paradigms and Scientific Revolutions

A revolution occurs when one paradigm replaces another. As normal science progresses, its puzzles succumb to the application of, or in Kuhn's words "the articulation" of, the paradigm. A small number of puzzles continue to be recalcitrant: unexpected phenomena that the paradigm cannot explain, phenomena that the paradigm leads us to expect but that don't turn up, discrepancies in the data beyond the margins of error, or major incompatibilities with other paradigms. In each case, there is a rational explanation for these anomalies within normal science; and often enough further work turns an anomaly into a solved puzzle. Revolutions occur when one of these anomalies resists solution long enough, while other anomalies succumb, to produce a crisis. As more and more scientists attach more importance to the problem, the entire discipline's research program begins to be focused around the unsolved anomaly. Initially small numbers of especially younger scientists without heavy investment in the ruling paradigm may cast about for a radical solution to the problem that the anomaly poses. This usually happens when a paradigm has become so successful that few interesting puzzles are left to solve. More and more of the younger scientists, especially, with ambitions and names to make, decide to attach more importance to the remaining unsolved puzzle. Sometimes, a scientist will decide that what could reasonably be treated as experimental error is something entirely new and potentially paradigm-wrecking. If the ultimate result is a new paradigm, what the scientist has done is retrospectively labeled a new discovery. When Roentgen first produced X-rays, he treated the result as contamination of photographic plates. The same plates became evidence of a significant phenomenon once paradigm shift had allowed for it. If the ultimate result is not incorporated by a paradigm shift, it gets treated as error—polywater for example—or worse, it may be thought of as fraud, such as cold fusion.

In developing a new paradigm, revolutionaries are not behaving in the most demonstrably rational way; nor are their—usually elderly—establishment opponents who defend the ruling paradigm against their approach acting irrationally. During these periods of crisis, when debate in a discipline begins to focus inordinately on the anomaly, neither side can be said to be acting irrationally. Defenders of the old paradigm have the weight of all its scientific successes to support their commitment. Proponents of the new one have only at most its solution to the anomaly that is recalcitrant to previous approaches.

Note that during these periods of competition between old and new paradigms, nothing between them can be settled by observation or experiment. This is for several reasons. To begin with, often there is little or no difference between the competing paradigms when it comes to predictive accuracy. Ptolemaic geocentric astronomy (with its epicycles) was predictively as powerful, and no more mathematically intractable, than its Copernican heliocentric rival. Moreover, "observational" data is already theoretically charged. It does not constitute an unbiased court of last resort. For Kuhn there is in the end no evidentiary court that will decide between competing paradigms which it is more rational to

embrace, which is closer to the truth, which constitutes scientific progress. This is where the radical impact of Kuhn's doctrine becomes clear.

A persistently unsolved and paradigmatically important anomaly will result in a scientific revolution only when another paradigm appears that can at least absorb the anomaly as a mere puzzle (and gives some promise of being able to explain the lion's share of the puzzles that the previous paradigm has already accommodated). In the absence of an alternative paradigm, a scientific discipline will continue to embrace its received one. But the grip of the paradigm on scientists is weakened; some among them begin to cast around for new mechanisms, new rules of research, new equipment, and new theories to explain the relevance of the novelties to the discipline. Usually in this "crisis situation," normal science triumphs; the anomaly turns out to be a puzzle after all, or else it gets set aside as a problem for the long-term future. Revolutions occur when a new paradigm emerges which disagrees radically with its predecessor. Sometimes new paradigms are advanced by scientists who do not realize their incompatibility with ruling ones. For instance, Maxwell supposed that his electromagnetic theory was compatible with the absolute space of Newtonian mechanics, when in fact Einstein showed that electrodynamics requires the relativity of spatio-temporal relations. But the new paradigm must be radically different from its predecessor insofar as it can treat as a mere puzzle what the previous one found an increasingly embarrassing recalcitrant anomaly. Paradigms are so all-encompassing, and the difference between paradigms is so radical, that Kuhn writes that scientists embracing differing paradigms find themselves in some sense in different worlds—the Aristotelian world versus the Newtonian one, the Newtonian world versus the quantum realm. Paradigms are, in Kuhn's words, "**incommensurable**" with one another. Kuhn took the word from geometry, where it refers to the fact that, for instance, the radius of a circle is not a "rational" fraction of its circumference, but is related to it by the irrational number pi. When we calculate the value of pi the result is never complete but always leaves a "remainder." Similarly, Kuhn held that paradigms are incommensurable: when one is invoked to explain or explain away another, it always leaves a remainder. But mathematical incommensurability is a metaphor. What is this remainder in the scientific realm?

According to Kuhn, although a new paradigm may solve the anomaly of its predecessor, it may leave unexplained phenomena that its predecessor success-fully dealt with or did not need to deal with. There is a trade-off in giving up the old paradigms for the new; an explanatory loss is incurred as the expense of the gain. For example, Newtonian mechanics cannot explain the mysterious "action at a distance" it required—the fact that gravity exerted its effects instant-aneously over infinite distances; this disturbing commitment is something the Aristotelian physics did not have to explain. In effect, "action at a distance"—how gravity is possible—became the anomaly that in part and after 250 years or so eventually undid Newtonian mechanics. But explanatory loss is not all there is to incommensurability. For even with some explanatory loss, there might yet be a net gain in explanatory scope for the new paradigm. Kuhn suggests that

incommensurability is something much stronger than this. He seems to argue that paradigms are incommensurable in the sense of not being translatable into one another, as poems in one language are often untranslatable. And this sort of radical incommensurability underwrites the further claim that paradigms do not improve on one another, and that therefore science does not accumulate in the direction of successive approximation to the truth. Thus the history of science is not unlike the history of art, literature, religion, politics, or culture; it is a story of change, but not over the long haul a story of "progress."

Kuhn challenges us to translate seventeenth-century phlogiston chemistry into Lavoisier's theories of oxidation and reduction. It cannot be done, without remainder. Perhaps you are inclined to say that phlogiston chemistry was all wrong, and needed to be replaced by a new paradigm. This is the sort of ahistorical approach to the nature of science that Kuhn condemned. After all, phlogiston chemistry was the best science of its day, it had a long record of success in solving puzzles, organizing instrumentation, and securing experimental support. And in the period before the heyday of phlogiston many scientists bent their genius towards alchemy. Isaac Newton was so devoted to the search for how to transmute lead into gold that he may have died of lead poisoning as a result of his many experiments. Are we to say that his mechanics was the greatest scientific achievement of a transcendent genius in physics while his alchemy was the pseudo-scientific mischief of a crackpot? Either we must condemn a century of scientific work as irrational superstition or design a philosophy of science that accepts alchemy and phlogiston chemistry as science with a capital "S." If phlogiston theory is good science, and cannot be incorporated into its successor, it is hard to see how the history of science can be a history of cumulative progress. It seems more a matter of replacement than reduction.

Reduction, recall, is the empiricist's analysis of the interrelation of theories to one another, both synchronically, in the way that chemistry is reducible to physics, and diachronically in the way that Newtonian seventeenth-century discoveries in mechanics are reducible to the twentieth century's special theory of relativity.

But does this reduction really obtain in the way that empiricism supposes? Kuhn explicitly denies that it does. And the reason is incommensurability. Reduction of the laws of one theory to the laws of a more basic theory requires that the terms of the two theories share the same meaning. Thus, the notions of space, time, and mass should be the same in Newton's theory and in Einstein's special theory of relativity, if the latter is just the more general case and the former is the special case, as reduction requires. The derivation of the laws of Newtonian mechanics from those of the special theory of relativity looks simple. All one requires is that "c," the speed of light, travels (like gravity) at infinite speed.

We illustrated this reduction by the use of Lorenz's length contraction equation in Chapter 8. There is a similar equation that appears to reflect the reduction of Newtonian mass to relativistic mass:

$$\text{MASS}_{\text{at rest relative to observer}} = \text{MASS}_{\text{in motion relative to observer}} \sqrt{1 - v^2\big/c^2}$$

When velocities are small compared to that of the velocity of light, the square root approaches 1. If the velocities approach zero, rest mass approaches mass in motion just as Newton requires.

The reason one requires this false but simplifying assumption to go from Newton to Einstein is that the special theory of relativity includes the Lorenz equation, and it tells us that the mass of an object varies as the ratio of its velocity to that of the speed of light with respect to an observer's frame of reference. Newton's theory tells us, however, that mass is conserved, and independent of relative or absolute velocity whether in proportion to the speed of light or not.

Though the two theories share the same word with the same symbol, m, do they share the same concept? Emphatically not. In Newtonian mechanics mass is an absolute, intrinsic, "monadic" property of matter, which can neither be created nor destroyed; it is not a relational property that chunks of matter share with other things, like "is bigger than." In Einstein's theory, mass is a complex "disguised" relation between the magnitude of the speed of light, a chunk of matter, and a location or "reference frame" from which the velocity of the chunk is measured; it can be converted to energy (recall $e = mc^2$). The change in the meaning of the word "mass" between these two theories reflects a complete transformation in world-view, a classic "paradigm shift." Once we as historians and philosophers of science see the difference between the meanings of crucial terms in the two theories, and discover that there is no common vocabulary—either observational or theoretical—that they share, the incommensurability between them becomes clearer. But, the physicist is inclined to say, "look here, the way we teach the special theory of relativity in the textbooks is by first teaching Newton's theory and then showing it's a special case via the Lorenz transformations. It is after all a case of reduction. Einstein was standing on the shoulders of Newton, and special relativity reflects the cumulative progress of science from the special case to the more general one."

To this Kuhn has two replies. First, what is reduced is not Newton's theory, but what we—in the thrall of the post-Newtonian, Einsteinian paradigm—imagine is Newton's theory. To prove otherwise requires a translation that would inevitably attribute incompatible properties to mass. Second, it is essential to the success of normal science that once it is up and running, it rewrites the history of previous science to make it appear just another step in the long-term continuous progress of science to cumulative knowledge of everything. The success of normal science requires the disciplining of scientists not to continually challenge the paradigm, but to articulate it in the solution of puzzles. Science would not show the pattern of accumulation that normal science exhibits without this discipline. One way to enforce this is to rewrite the textbooks to make it appear as much as possible that what went before today's paradigm is part of an inevitable history of progress that leads up to it. Whence the invisibility of previous paradigms, and the empiricist's blindness to what the history of science really teaches. For the

empiricist, understanding science comes from its contemporary textbooks and their "potted" history.

According to Kuhn we must take seriously the notion that scientific revolutions really are changes of world-view. The crucial shift from Aristotle to Newton was not the discovery of "gravity." It was in part the apparently slight change from viewing the distinction between rest and motion to be given by the difference between zero and non-zero velocity to viewing the difference to be between zero and non-zero acceleration. The Aristotelian sees a body moving at constant velocity as under the influence of a force, "impetus" they called it. The Newtonian sees the body as being at rest, under the influence of no (net) forces. The Aristotelian sees the swinging pendulum bob as struggling against constraining forces. The Newtonian sees the pendulum as in equilibrium, at rest. There is no way to express the notion of "impetus" in Newton's theory, just as there is no way to express Einsteinian mass in Newton's theory. More broadly, Aristotelian science views the universe as one in which things have purposes, functions, roles to play; Newtonian mechanics bans all such "teleological," goal-directed processes in favor of the interaction of mindless particles whose position and momentum at any time together with the laws of nature determine their position and momentum at all other times.

Because a new paradigm is literally a change in world-view, and at least figuratively a change in the world in which the scientist lives, it is often too great a shift for well-established scientists. These scientists, wedded to the old paradigm, will not just resist the shift to the new one, they will be unable to make the shift; what is more, their refusal will be rationally defensible. Or at any rate, arguments against their view will be question-begging because they will presume a new paradigm that they do not accept. To some extent we have already recognized the difficulty of falsifying a theory, owing to the underdetermination problem discussed in Chapter 11. Because paradigms encompass much more than theories, it is relatively easier to accommodate what some might call falsifying experience when adjustments can be made not just in auxiliary hypotheses but across a vast range of the intellectual commitments that constitute a paradigm. What is more, recall that there is no neutral ground on which competing paradigms can be compared. Even if underdetermination of theory by evidence were not a problem, the observational findings on which empiricists admit differing theories are missing. When allegiance is transferred from one paradigm to another, the process is more like a religious conversion than a rational belief shift supported by relevant evidence. Old paradigms fade away as their proponents die off, leaving the proponents of the new paradigm in command of the field.

Progress is to be found in science, according to Kuhn, but like progress in evolution, it is always a matter of increasingly local adaptation. The Darwinian theory of natural selection tells us that over generations the random variations in traits are continuously filtered by the environment so as to produce increasing spread of increasingly adaptative variations across a species. But environments change, and one environment's adaptation—say, white coats in the Arctic—is

another environment's maladaptation—white coats in the temperate forest. So it is with science. During periods of normal science, there is progress as more and more puzzles succumb to solution. But revolutionary periods in science are like changes in the environment, which completely restructure the adaptive problems a paradigm must solve. In this respect, science shows the same sort of progress that other intellectual disciplines show. And this is not surprising, for among the morals that many draw from *The Structure of Scientific Revolutions* is the conclusion that science is pretty much like other disciplines, and can make no claims to epistemic superiority. Rather, we should view the succession of paradigms in the way we view changes in fashion in literature, music, art, and culture broadly. Or perhaps we should view competing paradigms the way we view alternative normative ideologies or political movements. When we come to assess the merits of these units of culture, progress in approximating to the truth is rarely an issue. So too for science: In one of the last pages of his book Kuhn writes, "We may, to be more precise, have to relinquish the notion, explicit or implicit, that changes of paradigm carry scientists and those who learn from them closer and closer to the truth" (*The Structure of Scientific Revolutions*, first edition, ch. 13, p. 170).

Are Scientific Research Programs Rational?

It is not surprising that many philosophers of science and scientists have been unhappy with an account of science such as Kuhn's, which denied its progress, its cumulativity, and its rationality. Indeed, Kuhn himself in later writings seemed to reject the radical interpretation of *The Structure of Scientific Revolutions* presented here, which has become predominant.

Among the philosophers of science who sought an account of scientific change that accorded it rationality, one of the most visible was Imre Lakatos, a protégé of Karl Popper. It will be useful to sketch Lakatos' account, which he called "the methodology of scientific research programs," both to illustrate how some philosophers of science responded to Kuhn, and how they missed the force of his radical critique of progress in science.

According to Lakatos, scientific theories are components of larger cognitive units: research programs. Research programs are something like Kuhn's paradigms. But unlike a paradigm, a research program consists of statements, propositions, and formulae, and does not also include artifacts, experimental devices or distinctive measuring equipment, associated philosophical theses and other non-descriptive items. First, there is the research program's hard core: a set of assumptions about the world that are constitutive of the program and cannot be surrendered without giving it up altogether. For example, the Newtonian research program's hard core includes the inverse square law of gravitational attraction, while Darwinism's includes something like the PNS identified in Chapter 6. Surrounding the hard core is what Lakatos called its "protective belt," a set of further claims of the theory that function as auxiliary hypotheses. On the one hand, these theories are needed to apply the components of the hard core to explanation and prediction, but on the other hand, they may be changed

to avoid treating components of the hard core as falsified by evidence. Darwin's own quite mistaken theory of heredity is a good example: it was surrendered without any harm to the research program of evolutionary biology. Mendel's theory was added to the protective belt, with important consequences for the hard core. Two further components of a research program are the positive and negative heuristics, which include methodological rules that guide changes in the protective belt and enjoin revision of the hard core. The positive heuristic of Newtonian mechanics will include the injunction expressed in the principle of sufficient reason: "Every event has a cause: search for it!" The negative heuristic will deny "action at a distance"—causation without spatiotemporal contact (except by gravity).

A research program can be progressive or degenerate. It is progressive if over time its theories enable the scientists employing it to make new predictions, or at least to accommodate data already known but not originally employed to formulate the hard core of the program. Honoring the influence of Popper, Lakatos recognized that new predictions arise when scientists respond to the falsification of a program's predictions, by making changes in the protective belt, the positive or negative heuristic. If these changes enable them to derive novel expectations that are then realized, the research program is vindicated as progressive. The discoveries of Neptune and Uranus are classic examples of novel predictions in the Newtonian research program. Responses to falsifications that merely preserve the hard core without consequent new vindicated predictions are stigmatized as *ad hoc*.

When a program ceases to generate new predictions, and/or persistently invokes *ad hoc* changes in the protective belt or elsewhere, it is said to have become degenerate. According to Lakatos the rationality of scientific change consists in this: Scientists persist in the articulation (to use a Kuhnian term) of a research program so long as it remains progressive. Once it has ceased to be so for long enough, scientists begin to challenge some or all of the components of the original hard core, thus creating a new research program, distinguished from the degenerating one by a different hard core. Disciplines are characterized by a succession of research programs: each such program moves from progressive to degenerate, and is replaced by a new more adequate one accommodating its predecessor's novel predictions. This is progress across research programs, on Lakatos' view, and not merely the succession of alternative paradigms that Kuhn's account of scientific change suggests. A discipline that evinces Lakatos' model of research programs, their internal development and their succession, is a science that proceeds in accordance with rational standards of belief change. By this standard, the natural sciences appear to be safe from reproach, while many research programs in social science, among them the ones Popper stigmatized as pseudoscience—Marxian dialectical materialism, Freud's psychodynamic theory—are probably degenerate.

Has Lakatos really provided an account of rational scientific change? Kuhn's followers will hold that as it stands the differences with Kuhn's account of scientific change are cosmetic, except where Lakatos simply begs the question against

Kuhn's arguments. Lakatos neither does nor can give us a litmus test for when it becomes unreasonable to cling to a degenerating research program, still less a measure that would enable scientists to rank programs for progressivity. It is easy for a historian of science such as Kuhn to identify research programs that degenerated for a long period of time while retaining the confidence of scientists, and then began again to be progressive. Without such a litmus test, clinging to a degenerating research program may not be stigmatized as irrational, just as Kuhn argued. One might even count as a single progressive research program in physics the tradition that extends from Aristotle to Einstein, or at least as one worthy of rational support despite its period of degeneration during the temporary heyday of Newton's occult force, gravity.

Lakatos' theory faces problems of counting up new or novel predictions to decide whether successive or competing research programs are in fact progressive. Of course a follower of Lakatos could try to deal with some of these problems. The litmus test of scientific progress is, according to Lakatos, novel prediction. But why is this? We can rule out one apparently attractive answer immediately: The goal of science is improvements in technological application, and novel prediction is the best means to attain it. Pretty plainly, many scientists, for example cosmologists and paleontologists, do not share this aim of technological application. Indeed some among them—for instance biologists—hardly ever seek novel predictions. To suggest that science as an institution embraces aims that are separate from those of individual scientists is not unreasonable, of course, but a reason needs to be given for the goal we do attribute to it. What is more, even if technological application were the goal of science, it is by no means clear that a single-minded focus on novel prediction is the sole or best means to attain it.

As noted, much of the actual history of science shows that research programs (which for a time may degenerate, failing to provide novel predictions) end up doing better by way of novel prediction than their temporarily progressive competitors. In so doing, they reveal that the role of novel predictions is not in fact as decisive among scientists as Lakatos' methodology requires. Consider the vicissitudes of the wave and particle theories of light. The theory that light is a particle degenerated badly in the nineteenth century, owing to the experiments of Fresnel. This scientist argued that if light is composed of waves that interfere and reinforce each other, then there should be a bright spot at the center of a spinning disk, and no such spot if light were composed of particles. No one had ever performed the experiment to see if such a bright spot exists. The confirmation of Fresnel's novel prediction was striking evidence for the progressivity of his theory and the degeneracy of the particle theory. Yet a hundred years later, the particle theory's claims were vindicated in the shape of the photon.

Of course, Lakatos' account of scientific change can accommodate this history, according rationality to those who stuck with the research program of the particle theory throughout its period of degeneracy. But that is part of the problem. It is too easy for his account to do so. Whether a scientists sticks with a research program or not, it can still be defended as rational. Another problem

is that it was not owing to the search for any technological payoff that Fresnel made his novel prediction; indeed there was none of any importance.

So, why did Fresnel seek this novel prediction, and why did it have the effect of eclipsing the particle theory's research program for the better part of a century? Here is another attractive answer that we will have difficulty accepting: Science seeks theories with greater empirical content, and research programs whose theories make novel predictions that are vindicated have greater empirical content than ones that fail to do so. To begin with, this claim must be understood as one not about novel prediction in general, but the novel prediction of observable phenomena. Otherwise we are not talking about empirical content, but something else (theoretical content, whatever that is). This requires a controversial distinction between scientific vocabularies that are observational and those that are theoretical, which Kuhn would reject. It also requires a way of comparing theories for empirical content. But as we will see in the next chapter, distinguishing the *empirical content* of a theory from its logical, syntactical, mathematical, or other sort of *non-empirical form* is far from an easy task. What is worse, if theory choice is underdetermined by observation, as the previous chapter suggests, then it is evident that there can be competing or for that matter successive research programs, or at least theories, in a discipline with equal empirical content. Yet we never see the proliferation of such research programs or theories in the history of the sciences once the disciplines emerge from what Kuhn called the "preparadigm period." It must be that something else, then, is determining theory choice. This is, so to speak, where we came in on the story, of course. For we ended Chapter 11 searching for what determined the actual history of theory, research program, or paradigm choice in light of the fact that observation apparently does not suffice to do it.

Suppose we "privilege" novel prediction as the means to scientific progress by articulating an epistemological theory, an account of what knowledge consists in, that makes novel prediction a (perhaps uniquely) reliable indicator of justified true belief. Then of course insofar as science seeks knowledge, vindicated prediction will be the means to attain it, and a scientific discipline's attachment to novel predictions will turn out to be rational. Why, a proponent of Kuhn will ask, should we "buy into" this epistemology? Kuhn would argue that such an epistemology is not a "first philosophy" prior to science, that can stand in judgment of its rationality and epistemic progress. It is part and parcel of a paradigm. If so, embracing it without argument will simply beg the question against other epistemologies. But if a paradigm-free philosophy is impossible, as Kuhn holds, there is no neutral point of view from which to adjudicate competing epistemologies prior to applying them to assess the rationality of science. Of course, Kuhn would decline to divide paradigms up into cores, belts, and heuristics, each separately identifiable and changeable without impact on the other parts. Indeed, according to Kuhn, the central place of predictive novelty, especially in Newtonian science, is like the philosophy of logical positivism as a whole. Both are devices with which the Newtonian paradigm is defended.

Lakatos' methodology of scientific research programs will not provide the assurance we seek that, despite Kuhn's historical evidence, science is after all cumulative, progressive, or even rational. Thus, in this chapter, the stakes of the bet about the rationality of induction made at the end of Chapter 11 have been raised even higher. At the end of that chapter we faced the problem that the succession of scientific theories was not fully justified or for that matter explained by their relationship to observational evidence that is widely supposed to vindicate them. Now we are faced with the prospect that in addition to not being controlled by data, whatever it is that does control the course of science may not even be rational.

Summary

According to Kuhn, the unit of scientific thought and action is the paradigm, not the theory. Specifying what a paradigm is may be difficult, for it includes not just textbook presentations of theory, but exemplary problem solutions, standard equipment, a methodology, and usually even a philosophy. Among the important paradigms of the history of science have been the Aristotelian, the Ptolemaic, and the Newtonian in physics. Chemistry before Lavoisier, and biology before Darwin, were "preparadigm" disciplines, not yet really "scientific," for without the paradigm there is no "normal science" to accumulate information that illuminates the paradigm. The paradigm controls what counts as data relevant to testing hypotheses. There is, Kuhn argued, along with other opponents of empiricism, no observational vocabulary, no court of final authority in experience. Experience comes to us already laden with theory.

Crisis emerges for a paradigm when a puzzle cannot be solved, and begins to be treated like an anomaly. When the anomaly begins to occupy most of the attention of the figures at the research frontier of the discipline, it is ripe for revolution. The revolution consists of a new paradigm that solves the anomaly, but not necessarily while preserving the gains of the previous paradigm. What the old paradigm explained, the new one may fail to explain, or even to recognize. Whence it follows that scientific change—the succession of paradigms—need not be a progressive change in the direction of successive approximation to the truth.

Observation does not control inquiry, rather inquiry is controlled by scientists, articulating the paradigm, enforcing its discipline, assuring their own places in its establishment, except at those crucial moments in the history of science when things become unstuck and a revolution ensues—a revolution that we should understand as more in the nature of a palace coup than the overthrow of an old theory by one rationally certifiable as better or more correct.

Kuhn's *Structure of Scientific Revolutions* was the single most cited work in the humanities in the ten-year period from 1970 onward. This should be no surprise. Besides reorienting the philosophy of science towards the history of science, it provided potent ammunition for those humanists and others eager to put a stop to what they considered the imperialist or hegemonic pretensions of science.

Study Questions

1. According to Kuhn, to be successful, normal science must be authoritarian. Why does Kuhn make this claim and does it constitute a moral deficiency of science?
2. Defend or criticize: "The history of science is a history of errors by a fallible discipline. We can locate the defects of the paradigms that succeeded each other from Aristotle to Newton, from the point of view granted to us by the progress science has made."
3. Does Kuhn's approach to the nature of scientific change require that he at least have a paradigm-free vantage point?
4. What disanalogies are there between biological evolution by natural selection and scientific change?
5. How would Kuhn reply to charges that his account of science makes it look like just another religion?
6. Apply Lakatos' methodology of scientific research programs to one or another of the social sciences, identify the hard core, protective belt, positive and negative heuristic. Don't infer that if you have located them, the research program must be a "scientific" one. Why not?

Suggested Readings

Besides *The Structure of Scientific Revolutions*, Kuhn wrote important histories of science, including *The Copernican Revolution* and *Black-Body Theory and the Quantum Discontinuity*. In *The Essential Tension* Kuhn responds to some interpretations of *The Structure of Scientific Revolutions*. Another valuable source for understanding Kuhn's own later thinking about this book is *The Road since Structure*, which contains an autobiographical interview, edited by Conant and Haugeland. Thomas Nickles' anthology of papers by philosophers of science about Kuhn, *Thomas Kuhn*, is strongly recommended.

A. Bird, *Thomas Kuhn*, is a reliable introduction to Kuhn's views and influence. K. B. Wray, *Kuhn's Evolutionary Social Epistemology*, is a more advanced treatment.

An extract from *The Structure of Scientific Revolutions* is reprinted in Curd and Cover, along with important critical papers by McMullin, Longino, and Laudan.

Lakatos develops his account of scientific change in "Falsification and the Methodology of Scientific Research Programs," in a work containing several important papers on Kuhn's books, Lakatos and Musgrave, *Criticism and the Growth of Knowledge*. Another important post-Kuhnian account of scientific change, which is highly sensitive to the problems Lakatos' account faces, is Larry Laudan, *Progress and Its Problems*.

References at the end of Chapter 14 also bear on the subsequent debates about the relationship between the history and the sociology of science.

13 Naturalism in the Philosophy of Science

Overview

Much of the philosophical underpinnings for views like Kuhn's can be found in the work of an equally influential philosopher, W. V. O. Quine, who attacked logical positivism "from within" so to speak. A student of the positivists, Quine was among the first to see that the epistemology underlying their philosophy of science could not satisfy its own requirements for objective knowledge, and was based on a series of unsupportable distinctions. By casting doubt on the foundations of a tradition in philosophy that went back to Locke, Berkeley, and Hume, Quine made it impossible for philosophers of science to ignore the controversial claims of Kuhn and those sociologists, psychologists, and historians ready to employ his insights to uncover the status of science as a "sacred cow." But more important, this freed philosophers to make use of science in framing their metaphysics, epistemology, and their own philosophy of science.

Quine and the Surrender of First Philosophy

The Structure of Scientific Revolutions was published in 1962. The impact of its doctrines outside the philosophy of science is difficult to overstate. Kuhn's doctrine became the lever with which historians, psychologists, sociologists, dissenting philosophers, scientists, politicians, and humanists of every stripe

sought to undermine logical positivism's claims about science as objective knowledge controlled by experience. Meanwhile, within philosophy of science, developments that began earlier in the 1950s were reinforcing Kuhn's impact on logical positivism, if not on science itself. These developments owe a great deal to the work of philosopher W. V. O. Quine, whose thought provided some of the philosophical foundations often held to support Kuhn's historical conclusions.

Quine's two most influential works were a paper written in the early 1950s, "Two Dogmas of Empiricism," and a book written a decade later, *Word and Object*. Together these two works constituted a frontal attack on the epistemological assumptions shared by empiricists and rationalists since before Kant. These attacks were effective because they were "internal" to the shared philosophical tradition. They showed that critical assumptions of both epistemologies violated the standards for argument set by the empiricists and rationalists. The first dogma attacked in Quine's paper was the distinction between analytic and synthetic statements that Kant had introduced, and which Hume before him had adopted (under different labels). Recall from Chapter 2 that analytic statements are true in virtue of their defined meaning, while synthetic statements are true in virtue of facts about the world. The second dogma is the distinction between the empirical content and logical form of statements. Following empiricists like Hume, the positivists assumed this distinction in setting out their demand that meaningful statements have empirical content and their supposition that empirical content gave meaning. In *Word and Object* Quine elaborated his arguments against these two dogmas and set out to erect an alternative "naturalistic" epistemology and metaphysics that would avoid the problems for science that these two dogmas created. It was Quine's view that many of the epistemological and metaphysical problems that faced philosophers of science from Descartes to the present were the result of uncritically adopting these two "dogmas." Quine's work is what finally put an end to the research program of logical positivism. His thinking had a major influence on Kuhn and, together with Kuhn's book, set the agenda of the philosophy of science for the next few generations. Because of their influence on the philosophy of science it is worth setting out Quine's approach and alternatives to rationalism and empiricism in a little detail.

The traditional objectives of the philosophy of science were to justify science's claims to objective knowledge and explain its record of empirical success. The explanatory project of the philosophy of science is to identify the distinctive methods that the sciences share that enables them to secure knowledge; the justificatory project consisted in showing that this method is the right one, providing its foundations in logic—both inductive and deductive—and in epistemology, whether empiricist, rationalist, or some third alternative. These ongoing projects came up against traditional philosophical problems. In particular the underdetermination of theoretical knowledge by observational knowledge has made both the explanatory task and the justificatory one far more difficult. If observations underdetermine theory, then discovering the actual inference rules—the methods—that in fact are employed by science is a complicated matter that will require more than armchair logical theorizing. Philosophy will

have to surrender exclusive domain over the explanatory task (if it ever had such domain), to psychologists, historians, and others equipped to explore the cognitive processes that take scientists from hypotheses to data and back to theory. More radical has been the effect of underdetermination on the justificatory program. Underdetermination of theory by data means that no single hypothesis is supported or disconfirmed by any amount of observation. If data supports theory at all it does so in larger units than a single hypothesis. So it was that empiricist philosophers of science were driven to "**holism**" about justification: The unit of empirical support is the entire theory—both the hypothesis directly under test, every other part of the theory that supports the tested hypothesis, and all the auxiliary hypotheses needed to deliver the test.

Even more radically, the traditional philosophical gulf between justification and explanation came to be challenged by philosophers themselves. Explanations, as we noted in Chapter 1, cite causes, but causal claims are contingent, not necessary truths. The world could have been otherwise arranged and the laws of nature might have been different. That is why we need to undertake factual inquiry, not logical analysis, to uncover causes and provide explanations. Justification, however, is not a casual but a logical relationship between things. What may cause you to believe something does not thereby constitute evidence that supports your belief as well justified. Observing one thing may cause you to believe something, but it won't justify that belief unless there is the right sort of logical relation between them. These logical relations are studied naturally enough by philosophers, who seek their grounds: What makes the rules of logic—deductive or inductive—the right rules for justifying conclusions derived from premises, i.e. from evidence? The traditional philosophical answer to this question is that they are necessary truths that could not be otherwise.

Empiricists have difficulty with this answer because they hold that knowledge is justified by experience but that experience cannot demonstrate necessity. Therefore, logical principles that are to justify reasoning are at risk of being ungrounded themselves. For at least 200 years the empiricist's solution to the problem has been to treat all necessary truths, whether in logic or mathematics, as true by definition—as reports about the meaning of words, conventions we adopt to communicate. As such these statements are true by stipulation. The logical rule tells us that all inferences of the form

> if p then q
> p
> therefore
> q

are true because they reflect the meanings of the terms "if," "then," "therefore." Similarly, all the truths of mathematics, from $2 + 2 = 4$ to the Pythagorean theorem to Fermat's last theorem (there are no positive integer values of n greater than 2 such that $x^n + y^n = z^n$), are simply logically deduced from premises that are themselves definitions.

But twentieth-century work in the foundations of mathematics showed that mathematics cannot simply be composed of definitions and the consequences of them. When it was proved by Kurt Gödel that no set of mathematical statements can be both complete (enabling us to derive all the truths of arithmetic) and consistent (including no contradictions), the empiricist claim that necessary truths were all definitions came undone. We noted this outcome in Chapter 2. The reverberations from math to logic were inevitable. Empiricism needed a new theory of necessary truths, or it needed to deny that there were any. This is where holism and underdetermination re-enter the story.

A necessary truth, whether trivially true, like "All bachelors are unmarried," or less obviously true, like "The internal angles of a triangle equal 180 degrees," is one that cannot be disconfirmed by experience. But holism teaches us that the same can be said for statements that we consider to be contingent truths about the world, like "The spin angular momentum of an electron is quantized," or "The speed of light is the same in all reference frames." Scientists always prefer to make adjustments elsewhere rather than give these statements up. If holism is right, we can always preserve statements like these as true "come what may" simply by revising some other part of our system of beliefs about the world. But then, what does the difference between necessary truths and those contingent truths that we are unwilling to surrender come to? Necessary truths are true in virtue of the meaning of the words that express them, and contingent ones are true in virtue of facts about the world. But if two statements are both unrevisable, how can we tell whether one is protected from revision because of meanings and the other because of beliefs about the world? Notice this is an empiricist challenge to an empiricist thesis, or as Quine put it, a "dogma": that we can distinguish truth in virtue of meanings from truth in virtue of facts.

What are meanings? Recall the empiricist theory sketched in Chapter 8, which holds that meanings are ultimately a matter of sensory experience: The meaning of a word is given by definition in terms of some basement level of words that name sensory qualities—colors, shapes, smells, textures, etc. This theory of language resonates with our pre-philosophical belief that words name images or ideas in the head. But as we have seen it cannot make sense of the meaning of many terms in theoretical science. What is more, it has been hard to see how we could tell the difference between a truth about sensations and a sentence that reports a fact about the world: Suppose we define salty thus: "Salty is the taste one gets under standard conditions from sea water." What is the difference between this sentence and "Salty is the taste one gets under standard conditions from dissolved potassium chloride"? One cannot say the former is true in virtue of meaning, because it is meaning that we are trying to elucidate empirically by contrasting these two sentences. One cannot say that "potassium chloride" is a theoretical term, because "sea water" is equally not a label we can pin on a sample of clear liquid by mere visual inspection. We have to add the "standard conditions" clause to both sentences, because without them they would both be false (an anesthetized tongue won't taste either as salty). But having added the clause, both can be maintained as true, come what may in our experience. In

short, the meaning of words is not given by the sensory data we associate with them. Or if it is given by sensory experience, the relation is very complex. The conclusion that Quine came to was that "meanings" were suspect and no self-respecting empiricist philosopher should want to trade in them. A conclusion with wider support in the philosophy of science was "holism about meaning," a doctrine similar to and mutually supportive of the epistemological thesis of holism in the way that data tests theory.

If there are no meanings, or no truths of meaning distinct from truths about the world (if theory meets data as a whole and the meaning of a theory's terms is given by their place or role in a theory), then we have not just a philosophical explanation for underdetermination, but a philosophical foundation for incommensurability as well. Or at least we will if we part company with Quine in one respect. Despite his rejection of the empiricist theories of meaning and evidence, Quine did not surrender his commitment to an observational language with a special role in adjudicating competing scientific theories.

Given a continuing role for observation, we may not be able to compare theories sentence by sentence or translate competing theories into statements about what exactly we will observe under mutually agreed-upon circumstances. But we will be able rationally to choose between theories on the basis of their all-around powers to systematize and predict observations. The result for Quine and his followers was a sort of pragmatism that retained for science its claim to objectivity.

However, the implications of Quine's critique of empiricism's theory of meaning and evidence make for a more radical holism about mathematics, all the empirical sciences, and even philosophy for that matter. If we cannot distinguish between statements that are true in virtue of meaning and statements that are true in virtue of facts about the world, then there is no distinction in kind between the formal sciences, like mathematics, and the empirical sciences, such as physics or biology! Traditionally, mathematics and its branches—geometry, algebra, and logic—were held to be necessary truths. In epistemology, empiricists differed from rationalists about our knowledge of these necessities. Empiricists held them to be truths of meaning without content; this is why they are necessary, because they reflect our decisions about how to use the concepts of mathematics. Rationalists held that these truths were not empty or trivial disguised definitions and their consequences, but truths that experience could not justify. Rationalism could not in the end provide a satisfactory account of how we acquire such knowledge and so went into eclipse, at least as the basis for a viable philosophy of mathematics and science. But, to the extent that empiricism could not draw an empirically well-grounded distinction between truth in virtue of meaning and truth in virtue of facts about the world, its account of how we have knowledge of necessary truths collapsed. Quine's conclusion is that all statements we take to be true are of one kind—that there is no grounded distinction between necessary truths and contingent ones. So, mathematical truths simply turn out to be the most central and relatively unrevisable of our scientific hypotheses. Holism indeed.

What goes for mathematics, goes for philosophy too—including metaphysics, epistemology, logic, and the study of scientific methodology. On Quine's view, theories in these compartments of philosophy turn out also to be no different from theoretical claims in the sciences. A theory of the nature, extent, and justification of knowledge will turn out for Quine to be a compartment of psychology; metaphysics—the study of the basic categories of nature—will turn out to be continuous with physics and the other sciences, and its best theory will be the one that (when put together with what we know from the rest of science) gives us the most adequate account of the world, judged as a whole by its ability to explain and predict our observations. Methodology and logic also are inquiries to be pursued together with and not as independent foundations for the rest of science. Those methods and logical principles are most well supported which are reflected in the pursuit of successful science. Here the notion of "empirical adequacy" that we met in Chapter 2 is relevant. Quine's criterion for theory choice in philosophy and in science is empirical adequacy.

Instrumentalists argue for their doctrine from the privileged position of a prior philosophical theory, adherence to strict empiricism. Quine rejects the claim that there is some body of knowledge, say, a philosophy or an epistemology, that has greater credibility than science, and might provide a foundation for it. Though he holds that science should aim for empirical adequacy, he does so because this is the criterion of adequacy that science sets itself; what is more, unlike the instrumentalist, and like the scientist, Quine takes the theoretical claims of science about unobservables not just literally but as among the most well founded of our beliefs, because in the package of our beliefs that we call science, these are among the most central, secure, and relatively unrevisable. In fact, for Quine and his followers, science is as much a guide to philosophy as philosophy is to science. The difference between science and philosophy is one of degree of generality and abstractness, not a difference between necessary truths and factually contingent ones.

Naturalism, Multiple Realizability, and Supervenience

The resulting philosophy of science from these insights has come be called "**naturalism**." Among many philosophers, naturalism became the successor to empiricism largely as a result of Quine's influence. The label "naturalist" is one that many philosophers of science subsequently adopted despite differences among their philosophies.

As Quine defended it, naturalism's chief tenets are, first, the rejection of philosophy as the foundation for science, the arbiter of its methods, or the determinant of its nature and limits; second, the relevance of science to the solution of philosophical problems; third, the special credibility of physics as among the most secure and well-founded portion of human knowledge; and fourth, the relevance of certain scientific theories as of particular importance to advancing our philosophical understanding, in particular the Darwinian theory of natural selection.

Naturalists infer from the primacy of physics that there is only one sort of stuff in the world, physical matter or perhaps physical fields, or both, depending on how physics eventually settles the question of what is the basic stuff that the universe is composed of. This in turn arguably commits the naturalist to at least some minimal scientific realism, since taking physics at its literal truth requires one to believe that both the unobservable things it reports and the relationships it reports between them actually exist and are at least approximately as described in physics.

Second, naturalists will be committed to a denial of "dualism"—the thesis that the mind is a separate and distinct thing from the body (so that there are two basic kinds of things in the universe, whence "dual-ism"). This, however, immediately raises for naturalists the problem of showing how mental states, events, and processes can be physical ones, in the face of arguments against this possibility that go back to Descartes and before him. Indeed, it raises a broader problem of reconciling the monistic thesis of naturalism that there are only physical facts and things with the inability to reduce or fully explain all non-physical facts in terms of physics. It is not just the facts of psychology that cannot be reduced to those of physics. None of the significant facts described by any of the special sciences—i.e. the social and behavioral ones, or for that matter by the biological sciences—can be reduced to the laws of physics. Unless naturalism can explain away this fact consistent with its commitment to physics as the basic science, it will have to give up all these sciences. A scientifically inspired philosophy of science that cannot take any science seriously except physics is hardly a doctrine to be called naturalism, still less one that would attract serious support.

Recall from Chapter 8 the logical positivist thesis that the progress of science and the hierarchy of scientific disciplines are characterized by the reducibility or derivability of older narrower theories to newer, broader, and more accurate ones, and that the laws and theories of biology, psychology, sociology, economics, etc. should in turn be derived from the prior and more basic laws of biology, which should be derived from those of physics. We noted then that the problem with this thesis was that its requirement for **natural kinds** (those basic elements that represent the true categorization of the universe) cannot be defined in terms of more fundamental, lower-level theories. The few successes of physics (temperature = mean kinetic energy) are exceptions. Even the Mendelian gene cannot be defined in terms of the DNA double helix. Still less will it be possible to define the distinctive terminology of economics or even psychology in terms of neuroscience. So the naturalist faces a formidable problem in taking science seriously without jettisoning large parts of it.

The philosophically creative solution that naturalists have advanced to deal with this problem substantially enhanced the attractions of naturalism in the philosophy of science and in philosophy more broadly. What is more, since their solution exploits a particular scientific theory, it appears to substantiate naturalism's reliance on science and further strengthens its appeal.

First we need to draw a 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 of 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 16, except for isotopes).

The next step is to consider the apparently silly question: Can we "reduce" the type, kind, concept, "chair" to purely physical concepts, i.e. 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, etc. 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 "non-physical." No one has 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." Most naturalists hold that higher-level entities, such as chairs, are "nothing but" physical objects, even though the concept of a chair cannot be completely defined by appeal to concepts or terms drawn from physics. The higher-level entities "supervene" on lower-level ones in this specific sense: (a) Given any particular higher-level entity—say a particular chair, or for that matter a particular mind—it will have a particular physical composition (say a certain number of legs and arms, a seat and back perhaps, arranged in a certain way), or in the case of a mind, it will be a certain brain, with a certain specific set of connections among its neurons; (b) anything else whatsoever that has exactly the same material, physical composition, must also be respectively a chair with exactly the same functions, or a mind with exactly the same thoughts, feelings, and sensations. Naturalists are committed to the supervenience of higher-level phenomena on lower-level phenomena. But they are not committed to the derivational reduction of higher-level phenomena treated in the special sciences to phenomena described by physics. Supervenience, in short, does not imply reductionism. But why not?

Higher-level things that supervene on lower-level components or constituents are often "multiply realized." A chair can be made of wood or plastic or steel or straw or ice or … any of an indefinite number of different substances; it can have no legs or two or three or four or six … etc., it can be painted or

unpainted, have a cushion or arm rests, or … the list can go on forever. There are an infinite number of different combinations of physical, structural properties that could still be a chair. The same goes for any type, kind, concept, that we can define, describe, "characterize" in terms of its structure and its function. If it is a multiply realized type or kind, it will not be reducible to any finite list of structural components, even though its instances are composed of nothing but simpler components and ultimately physical ones. But why suppose this multiple realizability is very commonplace in nature?

Here is where science comes into the naturalist's explanation of why higher-level types can't be reduced to lower-level ones, even though physicalism is correct (i.e. even though everything is physical). The kinds and types of things at every level of organization that the sciences study bear the same sort of "multiple realization" relationship to things they are made up of that "chair" bears to the things chairs can be made up of. And this fact is itself almost always the result of a natural process that science has discovered: Darwinian natural selection.

Biological functions are what evolutionary biologists call adaptations. They are the results of a Darwinian process of blind variation and natural selection (environmental filtration), as we have seen in Chapter 9. Indeed, many philosophers have defined biological function in terms of Darwinian etiology of variation and selection. Such a process will select for any structure at all that happens to fulfill the same function. If the function that is selected for is camouflage or mimicry, then there may be selection for skin color, or shape, or ability to remain motionless, or any of a dozen other ways to solve this "design problem." If the function selected for is to transport oxygen from the lungs to the capillaries, then any of a dozen different hemoglobin molecules may do the job just as well. In short, selection for function is blind to structure: Any two or more structures that accomplish the same or even roughly similar functions will be selected for in nature.

In consequence, as more and more complicated levels of organization emerge through the persistent process of Darwinian natural selection, at every level there will be kinds or types of things with multiply realized structures. These kinds will not be reducible to the kinds at lower levels owing to multiple realizability. But this fact will be completely consistent with physicalism—the thesis that there are only physical things. Or at least the irreducibility of higher-level things will be consistent with physicalism if the kinds of events, processes, and things discovered in every science result from Darwinian processes, since these processes inevitably produce multiple realizability. One tip-off that this must be the case in all of the special sciences is that the "vocabulary"—the characteristic kinds and types—of all of these disciplines is functional. The social and behavioral sciences are not concerned with behavior except insofar as it appears to be purposive, goal-directed, and to show the ends/means economy that we associate with rational agents. If, as the naturalist holds, the only source of the appearance of purpose is the process that Darwin discovered, then this solution to the problem of irreducibility will be just what the naturalist needs.

Darwinian selection may well be responsible for all the apparent irreducibility of non-physical processes, including psychological ones, to purely physical ones, without having any untoward metaphysical consequences for physicalism (and naturalism) as a broader methodological and philosophical thesis.

The importance of Darwinian theory as a scientific guide to the solution of philosophical problems owes its account to how blind mechanistic processes can give rise to the *appearance* to us of purpose and design in a world of blind variation and natural selection. Recall the problem of teleological or goal-directed processes and their causal explanation, discussed in Chapter 1 and Chapter 6. Physical science has no conceptual room for final causes, for effects in the future bringing about causes in the past. Still less does it have scope for an omnipotent designer who brings things about to suit his or her desires. This is why the naturalistic world-view found Darwin's theory so attractive, because it provided a causal mechanism—the perpetual occurrence of variation (through mutation and recombination) in traits that just happened to be heritable. If we can use the same mechanism of random heritable variation and selection by the environment to explain other apparently purposive non-physical processes, especially human affairs, we will have accommodated these processes at least in principle to a single scientifically coherent world-view—a naturalistic philosophy.

Exploiting Darwinism, philosophers have sought to provide a naturalistic account of scientific change, similar in some respects to Kuhn's account of scientific progress as local adaptation. However, noticing that the environment—physical reality—that filters for improvements in scientific ideas does not change the way that the biological environment changes over time, some philosophers have tried to construct a Darwinian motivation for scientific realism in contrast to Kuhn's account of progress as only local. Others, following Laudan, have drawn attention to pessimistic induction to throw cold water on the notion that blind variation and natural selection of scientific theories should ensure successive approximation to the truth. This was certainly van Fraassen's line of argument. He invoked Darwin's theory to explain why theories improve over time in empirical adequacy—predictive power: They are selected for doing so by an environment that puts a premium on technological application, an environment created by *Homo sapiens*. By contrast, at least one other influential naturalist of the late twentieth century, Philip Kitcher, has argued that science shows cumulative progress in approximation to the truth at least in large measure because it is the product of a species—us—whose cognitive equipment has been selected for discovering significant truths about their environment. Indeed, natural selection has conferred such powers on human cognition that we need not take the underdetermination of science by evidence very seriously. We have been selected for entertaining only those alternative theories it is worth considering.

Other philosophers have appealed to natural selection to deal with problems about human cognition that have daunted philosophy and psychology since Descartes' time. Dualists following Descartes have argued that mental processes, especially those of which we are conscious, cannot be physical processes in the brain. It is clear that this view is impossible to reconcile with naturalism and so

it is a natural target for naturalistic philosophers, especially followers of Quine. Since the only resource that naturalistic philosophers and psychologists have for dealing with an apparently purposive process such as human thought is to show that it is really identical to the operation of some Darwinian process, the theory of natural selection has come to bear a great deal of weight in the philosophy of psychology.

Finally, naturalists look to Darwin's theory to help underwrite and explain the nature of human ethical norms and their grounds. Naturalists cannot appeal to anything above and beyond what science discloses to justify beliefs. Science as the sole source of justification for beliefs is the core commitment of naturalism. This means of course that when it comes to justifying moral claims and theories, naturalists are committed to finding a way to derive the normative statements about what ought to be the case in human conduct and social institutions from what is the case as a matter of biology.

This makes it clear why the philosophy of biology has become a growth area in the philosophy of science over the last half decade. Most human behavior appears to be purposive and goal-directed, to show the economy of means to ends that continues to make ordinary thought and common sense thoroughly teleological. In this way common sense shares much more with the Aristotelian world-view that characterized science until the Newtonian revolution of the seventeenth century. All subsequent development in physical science repudiated future causation and past purposes written into nature to guide it. This is what made Darwin's theory indispensable to the Newtonian world-view. Insofar as contemporary naturalists among philosophers and social scientists seek a non-teleological theory of human affairs that satisfies the prohibition of purpose, it has no alternative but a Darwinian approach to human phenomena.

Thus Quine's naturalism pushed the problems of the philosophy of biology near the top of the agenda of the philosophy of science. Some of these problems are familiar from our discussion of the theory of natural selection and its model-centric development in Chapter 9. For a relatively long time philosophers of biology have puzzled over the definition of fitness, the key term in evolutionary theory. The problem, recall, is defining the concept in terms of reproduction without making the resulting theory true by definition, or without empirical content. The Quinean denial that we can distinguish statements true by definition from those true in virtue of facts about the world obviously did not absolve philosophers of biology of the need to solve this problem. This was not least because creationists and others opposed to Darwinian theory sought to apply Popper's demand that real scientific theories be falsifiable to attack evolutionary theory as merely metaphysics and not real science. The resulting controversy brought many of the issues treated in Chapters 9 through 12 into public debate (and in the U.S. into courtrooms that adjudicated constitutional issues of the separation of church and state).

As a result of developments in biology in the last third of the twentieth century, Darwinian theory began to be applied in a broad range of social sciences to explain a variety of institutions, practices, norms, and attitudes as a result

of processes of blind variation and natural selection. Darwinian theory is in its origin a theory of hereditary transmission of genetically hard-wired traits. Accordingly the application of the theory was deemed by some opponents and some proponents to suggest that important human traits are genetically fixed and not open to environmental modification. This view appeared to many people to be morally dangerous. It appeared to excuse societies from taking steps to ameliorate human conditions on the ground that they were the result of traits fixed by genes, by "nature." It also suggested that long-lived social institutions were beneficial because they must have been selected for their contribution to fitness. This was another reason for politically, socially, and economically conservative policies. The role of Darwin's theory in the debate about these matters added to the incentives that philosophers acquired, along with a commitment to naturalism, to come to grips with the theory.

Naturalism's Problem of Justification

Naturalism leaves a major problem as yet unsolved. Recall the distinction between justification and causation. Justification gives grounds for the truth of belief; causation does not. Or at least so it seems. In the empiricist's hands, justification is a logical relation (employing deductive or inductive logic) between evidence (sensory experience) and conclusion, and logic is a matter of meaning. Naturalists, or at least Quineans, cannot help themselves to this way of drawing the distinction between causation and justification. Yet draw it they must. Without recourse to a "first philosophy"—some body of *a priori* truths or even definitions—naturalism can only appeal to the sciences themselves to understand the inference rules, methods of reasoning, methodologies of inquiry, and principles of epistemology that will distinguish between those conclusions justified by evidence and those not justified by it.

Now, suppose one asks of a principle of logic or methodology whether this method or rule is itself justified or well grounded. The empiricist has an answer to this question: The rule or method is necessarily true, and its necessity rests on our decision about how to use language. We may dispute this argument, and naturalists will do so, because it trades on notions in dispute between empiricists and naturalists—notions like "necessity" and "meaning." But what can naturalists say when asked to ground their own justificatory rules and methods? Appeal to a "first philosophy"—an epistemology prior to and more secure than science—is out of the question. And naturalism cannot appeal to science or its success to ground its rules without begging the question. The appeal to a "first philosophy" would be circular, and grounding its rules on science's technological success would be to surrender naturalism to a first philosophy—in this case, one called "pragmatism."

Naturalism justifies the epistemology, logic, and methodology it recommends because this trio of theories and rules emerges from successful science—i.e. research programs that provide knowledge about the way the world works. But if asked why they claim that successful science provides such justified conclusions, naturalists cannot then go on to cite the fact that successful science proceeds by

rules and methods that certify its conclusions as justified, because these rules and methods are themselves certified by science's success. Naturalism would be reasoning in a circle. This problem is particularly acute for Quine, because many of his arguments against empiricism's answers to these questions by appeal to concepts of logical necessity and meaning accused these answers of circular reasoning.

To appeal to the practical, technological, applied success of science might solve the naturalist's justificatory problem. But the result would no longer be naturalism. Science does in fact have a magnificent track record of technological application with practical, pragmatic success. But why should this provide a justification for its claims to constitute knowledge or its methods to count as an epistemology? It does so only if we erect a prior first philosophy, call it pragmatism, after the early twentieth-century American philosophers—William James, C. S. Peirce, and John Dewey—who explicitly adopted this view. This philosophy may have much to recommend it, but it is not naturalism, for it begins with a philosophical commitment prior to science, and may have to surrender those parts of science that are incompatible with it.

Naturalism is thus left with an as yet unfulfilled obligation. It aims to underwrite the objectivity of science, its status as ever-improving knowledge of the nature of things. It also aims to reflect the actual character of science in its philosophy of science, without giving either philosophy or history a privileged role in the foundations of science or the understanding of its claims about the world. Yet it needs to answer in a way consistent with its own principles the question of its own justification.

Summary

Much more than Kuhn, among philosophers it was W. V. O. Quine who unraveled the tapestry of logical positivist theories of science.

Quine began by undermining two distinctions: that between statements that are true as a matter of logic versus those that are true as a matter of content or empirically observable fact. It may be surprising but once this distinction (well known to philosophy since Kant) is surrendered, everything in epistemology and much in the philosophy of science becomes unstuck. The denial of this distinction gives rise to holism about how theory confronts experience, and to the underdetermination that seems to support Kuhn's approach to the nature of science. But it also gives rise to a stronger commitment to science by some philosophers than even to philosophy, or at least it gives rise to the idea that we must let contemporary science be our guide, instead of seeking science's foundations in philosophy. Philosophers, largely followers of Quine, who have adopted this view label themselves "naturalists," a term that others, especially sociologists adopting incompatible views, unfortunately have also adopted.

Naturally, neither Quine nor other philosophers are prepared to accept Kuhn's apparent subjectivism about science as the correct conclusion to draw from their attack on empiricism. This places on the table a new problem, beyond Hume's

problem of induction, of finding a foundation for science as objective knowledge consistent with Kuhn's arguments. The recent vicissitudes of work on this problem is the subject of the next chapter.

Study Questions

1. Quine claimed to be an empiricist. Did he have a right to this label for his views?
2. How would Hume or Kant, two philosophers with very different views, who embraced the analytic/synthetic distinction, respond to Quine's argument?
3. Quine once said "philosophy of science is philosophy enough." Give an interpretation of this claim that reflects Quine's claims about the relation between science and philosophy.
4. Is naturalism question-begging? That is, does according the findings of science over philosophical theorizing rest on mere assertion that science is our best guide to the nature of reality?
5. Mathematical truths appear to be *a priori*. What problems does this raise for naturalism? Can a naturalist be a Platonic realist?

Suggested Readings

Quine's attack on empiricism emerges in *From a Logical Point of View*, which contains his extremely influential essay, "Two Dogmas of Empiricism." This too is required reading for anyone interested in the philosophy of science. It is reprinted in Curd and Cover. Quine, *Word and Object* is a later work that deepens the attack on empiricism, and develops the doctrine of underdetermination that was so influential on Kuhn and others. Balashov and Rosenberg's anthology includes "Two Dogmas of Empiricism." Scott Soames, *Philosophical Analysis in the Twentieth Century*, especially volume 1, expounds and assesses Quine's work and influence.

Naturalism in the philosophy of science is expounded and defended in P. Kitcher, *The Advancement of Science*. Daniel Dennett's *Darwin's Dangerous Idea* is a magisterial introduction both to evolutionary theory and to its impact on philosophy in general. Fred Dretske, *Explaining Behavior*, approaches the mind–body problem from this perspective in an important but accessible work. More difficult but important is Jaegwon Kim, *Physicalism or Something Near Enough*.

Jerry Fodor introduced the problem of multiple realizability for reduction in "Special Sciences: Or the Disunity of Science as a Working Hypothesis," reprinted in Lange.

14 The Contested Character of Science

Overview

Kuhn's doctrines have generally been interpreted so as to give rise to **relativism**—the theory that there are no truths, or at least that nothing can be asserted to be true independent of some point of view, and that disagreements between points of view are irreconcilable. The result of course is to deprive science of a position of strength from which it can defend its findings as more justified than those of pseudoscience; it also undermines the claims of the so-called "hard sciences"—physics and chemistry—to greater authority for their findings, methods, standards of argument and explanation, and strictures on theory construction, than can be claimed by the "soft sciences" and the humanities. Postmodernists and deconstructionists took much support from a radical interpretation of Kuhn's doctrines, and from other fashionable philosophies, for the relativism they embraced.

Among sociologists of science especially, a "**strong program**" emerged to argue that the same factors that explain scientific successes must also explain scientific failures, and this deprives facts about the world—as reported in the results of observations and experiments—of their decisive role in explaining the success of science.

> These doctrines had a liberating effect on the social and behavioral sciences and other disciplines that had hitherto sought acceptance by aping "scientific methods" but no longer felt the need to do so. The sociological, and even more political, focus on science revealed its traditional associations with the middle class, with capitalism, its blindness towards the interests of women, and indifference to minorities. Philosophers of science, especially the feminists among them, have increasingly been sensitive to these facts about science's past and present, which has led to insights about how science should be pursued hereafter.

Methodological Anarchism

The interaction of the naturalism that Quine inspired, and the reading of the history of science that Kuhn provided, together had a profoundly unsettling impact on the philosophy of science. It shook centuries of philosophical confidence that it understood science. This sudden loss of confidence that we know what science is, whether it progresses and how it does so, and what the sources of its claims to objectivity can be, left an intellectual vacuum. It is a vacuum into which many sociologists, psychologists, political theorists, historians, and other social scientists were drawn. One result of the heated and highly visible controversy that emerged was to make it apparent that the solution to problems in the philosophy of science requires a re-examination of the most fundamental questions in other compartments of philosophy, including epistemology, metaphysics, the philosophy of language, and even portions of moral and political philosophy.

Kuhn held that paradigms are incommensurable. This means that they cannot be translated into one another, at least not completely and perhaps not at all; incommensurability also implies explanatory losses as well as gains, and no common measuring system to tell when the gains are greater than the losses. Incommensurability between paradigms reaches down to their observational vocabulary and deprives us of a neutral position from which to assess competing paradigms. The result is a picture of science not as the succession of more and more complete explanations of a wider and deeper range of phenomena, nor even the persistent expansion of predictive power and accuracy over the same range of phenomena. Rather, the history of science is more like the history of fashion, or political regimes, which succeed one another not because of their cognitive merits, but because of shifts in political power and social influence. This conception of the history of science is an invitation to epistemic relativism.

Ethical relativism is the claim that deciding which actions are morally right varies from culture to culture and there is no such thing as objective rightness. Ethical relativism is seen by its proponents as an open-minded and multicultural attitude of tolerance and understanding about ethnic differences. Ethical relativism leads inevitably to skepticism about whether there really is any such thing as absolute moral rightness at all. Epistemic relativism similarly

makes knowledge (and therefore truth) relative to a conceptual scheme, a point of view, or perspective. It denies that there can be an objective truth about the way the world is, independent of a paradigm, nor consequently any way to compare paradigms for truth, objectivity, or epistemic warrant. Kuhn was ambivalent about whether to plead guilty to the charge of epistemic relativism among paradigms.

But the situation may be even more fraught than Kuhn supposed. For there were philosophers and others eager to transform Kuhn's claims about the broadest paradigms that characterize century-long epochs of normal science into the incommensurability of individual scientific theories even within the ambit of normal science. And Quine's fundamental philosophical arguments gave them the resources to do so. Most influential among these philosophers was Paul Feyerabend. Adopting Kuhn's insights about the irreducibility of **Aristotelian mechanics** to Newton's theory, and of Newtonian mechanics to Einstein's, Feyerabend argued that the impossibility of translating the key concepts of impetus into inertia, or absolute mass into relative mass, reflects a barrier to reduction among all theories. The reason is the holism about meaning that Quine's insights spawned. The meaning of a theoretical term is not given by its connection, direct or indirect, to observation, because theory does not meet observation word by word or even sentence by sentence, but only as a whole. Thus meanings are theoretical. The meaning of a theoretical term is given by its place in the structure of the theory in which it figures. Change one or more parts of a theory and the result is not an improvement on the same theory, but an altogether new and different one. Why? Because the new theory is not about the same subject matter as the old theory, since its words have different meanings. "Electron," though it may be an inscription in Bohr's theory, Thomson's theory, Heisenberg's and Schrödinger's, no more means the same thing in each of them than does "cat" mean the same in "pussy cat," "catastrophe," "cool cat," and "cat-o'-nine-tails."

Denying this holistic claim about meanings requires an entire theory of meaning, or at least a reasoned objection to Quine's attack on meanings. When added to the denial of an observational language that could frame statements about data, which might enable us to choose between theories, the result is what Feyerabend praised as "methodological anarchy." He called it this because the result is that there is no cognitive basis to choose between theories. In particular, earlier and "well-established" theories have no claim to our adherence above later and less well-established ones. And Feyerabend praised this outcome because he held that such anarchy stimulates scientific originality and creativity. After all, if Newton had been required to advance a theory that could treat Aristotle's as a special case, or had Einstein been required to do so for Newton just because of the explanatory and predictive successes of Aristotle's or Newton's theory, neither Newton nor Einstein would have produced the great scientific revolutions that bear their names. Just as moral relativists think their insight is emancipatory and enlightened, so did Feyerabend think his epistemic relativism a good thing.

Feyerabend, and other relativists, would stigmatize naturalism from just this perspective. Like Kuhn, and all naturalists for that matter, relativists will

agree that an epistemology and methodology are parts of a paradigm, or in fact components of a theory, although perhaps these components are expressed grammatically in the imperative instead of the indicative. As such, epistemology and methodology don't provide an independent position from which to adjudicate scientific advance, or even the status of a discipline as "scientific" with a capital "S." These relativists would seize upon the problem of circularity that faces naturalism to substantiate their claim that any particular theory, paradigm, or discipline is but one among many "ways of knowing," and that there is no such thing as one of them being correct and the others mistaken. So far as the relativist is concerned, "anything goes." This in fact was the title of a book in which Feyerabend most forcefully argued for this view. Instead of a brief biography Feyerabend provided his astrological chart on the book's dust-jacket. He meant to suggest that astrology was as informative about the author as personal facts about his education, career, and previous books might have been.

The "Strong Program" in the Sociology of Scientific Knowledge

If one takes the position that from a philosophical point of view "anything goes," the question emerges, why has science taken the particular route that it has over time? (And why has science been so successful?) For the relativists the answer cannot be that the history of science is the history of inquiry "tracking the truth." According to the relativist, the way that the world is, independently of science, can have no role in determining the shape of any particular science or science in general. That is because there is *no way* that the world is, independent of how science views it at a particular time, and there is no objective way to know it, either. We can either take this claim literally or figuratively, as we will see. If the history of science is not explained by dispassionate study of the world by objective and disinterested scientists, it must, like all the history of all other social institutions, be the outcome of social, political, psychological, economic, and other "non-cognitive" factors. So, to understand science, the particular sciences, and the nature of scientific change, relativists argue, we must do social science. For example, to learn why Darwin's theory of evolution triumphed does not require that we understand the fossil record, still less the sources of variation and environmental filters. It requires that we understand the social and political forces that shaped theory construction and acceptance in the nineteenth century. Once we understand the ideological needs of nineteenth-century laissez-faire capitalism to justify relentless competition in which the less fit were ground under and progress was a matter of market competition, the emergence of the Darwinian paradigm should be no surprise. That the history of science should be rewritten by each successive paradigm is now understandable not just because normal science requires ideological discipline, but because political domination requires it as well.

The denial that tracking truth has a special role in the explanation of scientific change, which it lacks in, say, changes in literature or fashion, led in the 1980s to an important new movement in the sociological study of science, and

a concomitant claim by this movement that sociology must displace philosophy as our source for understanding science. The so-called "strong program" in the sociology of science set out to explain both scientific successes and failures on the same basis. Since what distinguishes those scientific developments that are accepted as advances from those rejected (with hindsight) as mistaken cannot be that the former reflect the way the world works and the latter do not, both must be explained in the same way. The "weak program" of trying to explain scientific failure as a function of social, political, and other non-cognitive forces that knocked it off track in the pursuit of truth must be replaced by the "strong program" of trying to explain all scientific theorizing—whether failed or successful—outside the realm of its alleged correspondence with reality. The sociologist David Bloor described this as the "symmetry thesis": It leaves no space for any argument that what explains successful scientific theorizing is that it is more rational than unsuccessful theorizing.

These sociologists and other social scientists sought to study the close details of scientific work, and concluded that like other social products, scientific agreement was "constructed" through "negotiation" between parties whose interests are not exclusively or perhaps even predominantly directed at describing the way the world works. Rather their interests are personal advancement, recognition, material reward, social status, and other benefits that bear no connection to the declared, publicly stated, advertised objectives of science: the disinterested pursuit of truth. In the hands of some radical students of science, the thesis that scientific findings are constructed becomes the claim that the world external to scientific theory, which realists identify as the independent reality that makes scientific claims true or false, is itself a construction without existence independent of the scientists who agree upon their descriptions of it. This "idealism," according to which to exist is nothing more than to be thought of, goes back in philosophy of science to the eighteenth-century philosopher George Berkeley, and certainly has the explicit support of at least some perhaps incautious remarks of Thomas Kuhn, which suggest that proponents of differing paradigms live in differing worlds.

Among the most prominent works of these sociologists was Bruno Latour and Stephen Woolgar's *Laboratory Life*, in which the authors inserted themselves in a molecular biology laboratory in much the same way that a cultural anthropologist might attempt to "go native" among a completely foreign people, immersed in a culture very different from the anthropologist's own society. Latour and Woolgar are examples of how a twentieth-century tradition in French intellectual life approached the study of science. Prior to their work, there had been a long tradition of impressive accomplishment in the history of science due to figures like Poincaré, Duhem, Bachelard, Cavaillès, and Granger. Philosophers who (long before Kuhn) took science's history with the utmost seriousness shifted the attention of English-speaking philosophers of science to the importance of its history without necessarily challenging the primacy of logical or rational understanding. In anthropology and sociology, however, a separate current arose in French social science that called itself "structuralism." Following

Claude Lévi-Strauss, it wedded the "thick"—i.e. detailed—description of the life of the "natives" that was common in both English-speaking and French cultural anthropology to the claim that social facts are autonomous from and irreducible to facts about individuals and their psychological states, and that they controlled these states and the actions that issue from them. Latour and Woolgar applied the combination of thick description and structural determination to the research lab. They sought to show two things: First, that the results of laboratory experiment do not speak for themselves, but are created, put together by discussion, dispute, and compromise governed by this structure and the roles it dictated. Knowledge is not discovered but constructed in accordance with social norms. Second, and closely related, that the winners of this negotiation are not the individuals with the best evidence, arguments, methods, or logic, but those who occupy the most exalted positions in a social structure that exists independent of any person, and shapes their behavior. Concepts like "truth," "evidence," "fact," "reality" are rhetorical tools wielded by individuals to win debates. Like proponents of the strong program, these anthropologists of laboratory life repudiate the suggestion that the notions philosophers take seriously make any real contact with any non-social, uninterpreted, independently fixed reality or nature that drives scientific results.

English-speaking social historians of science were never as sympathetic to the theoretical superstructure common to twentieth-century continental students of science, but they were equally out of sympathy with the shared assumptions of philosophers about how reality and rationality between them work together to create scientific achievements. Among the most influential of these were Stephen Shapin and Simon Schaffer. In *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*, they came to conclusions somewhat similar to Latour and Woolgar's about how scientists construct facts. They would have repudiated any overarching theoretical framework as governing their discoveries about science. But, others, especially philosophers of social science, would have had no trouble detecting a commitment if not to French structuralism, at least to the view that social facts and forces have functions for the maintenance of social peace and conflict resolution that individuals do not recognize, yet still govern their conduct.

As Shapin and Schaffer remind us, in the late seventeenth century Britain was recovering from a social upheaval that resulted in 20 years of strife, fundamental changes in religious orthodoxy, the overthrow of parliamentary prerogatives, the execution of a king, and the establishment of a military dictatorship. The re-establishment of a stable social order rested on the emergence of a means by which disputes that appeared to be abstract—for example, religious differences or constitutional disagreements—could be settled without force or violence. For otherwise, in the period after the Restoration, they would not even have been raised, discussed, or settled. Rather they would have had to be ignored as too dangerous to social peace. Insofar as seventeenth-century science was still inseparable from philosophy, and philosophy from religion, controversies surrounding scientific questions remained potentially dangerous no matter how slight their practical application.

In this context, Shapin and Schaffer argued, the emergence of experiments and experimental methods independent of theoretical and theologically controversial issues provided an avenue for scientific advance that would not threaten the social order. As such it soon acquired both scientific authority and political toleration. The scientists whom we esteem for having established experimental method may have thought they were uncovering a means of providing objective knowledge. Their successors and we ourselves may suppose that their fame and the persistence of the methods they established were based on the method's capacity to confirm and certify truths about nature. But this is wrong. In fact, the experimental method persisted because it filled an important social function for science. It ensured that science would not destabilize the social order and so could be safely allowed to survive and flourish. Shapin and Schaffer's study of the early history of the Royal Society of London in the seventeenth century sought not only to show how social facts, forces, and functions explain the emergence of the scientific revolution and the hegemony of the experimental, empirical method, but analyze how, without working scientists recognizing it, these processes reshaped fundamental concepts of what constituted facts and findings, their scientific authority over theory and explanation, in ways that fooled them, their successors, and even us today into thinking that science discovers (as opposed to constructs) a reality existing independent of us. For this illusion is one especially effective in maintaining its power, and that of its patrons.

Notice that these and other approaches to the sociology and anthropology of scientific knowledge cannot claim to provide knowledge about the real nature of science, without admitting and accepting that there really is such a thing as knowledge. They require therefore some minimal account of how beliefs can be true and justified, for that after all is what knowledge is—justified true beliefs about things, expressible in meaningful statements about them. For example, they have to accept the responsibility of testing the structural and functional theories of social institutions to which they are explicitly or implicitly committed. In doing so they must already accept a minimum common ground that the traditional philosophy of science has identified and tried to articulate. The social students of science so far treated in this chapter dissent mightily from the claims about science that are traditional among philosophers, but accept the epistemological and logical constraints that are required for any claim to constitute knowledge about the world or the sciences. The same may not be true for some of the views later to be discussed in this chapter.

Several of the conclusions advanced by participants in the social study of science have given encouragement to certain philosophies of social science and certain accounts of the nature of knowledge in the humanities as well. Thus, some qualitative social scientists came to defend their methods and results against attack from empirical and quantitative social scientists by claiming for these the status of a separate and incommensurable paradigm in which differing social forces operate in accordance with differing institutional rules to generate outcomes, theories, findings, and explanations that are no less "objective" (and no more "objective") than the outcome of natural science. These defenders of

qualitative social science went onto the counterattack, arguing that the empirical, quantitative, and experimental paradigm is incapable of dealing with human meaning, significance, and interpretation; that these are the essential dimensions along which human action, emotion, and value are to be understood; that the natural science paradigm cannot even accommodate the notion of semantic meaning, let alone human significance; and that the sterility and frustration of much social science is the result of slavishly attempting to implement an inappropriate paradigm from the natural sciences. The inability to surrender the quantitative paradigm in the face of anomalies of the sort that should lead to the questioning of normal science is a reflection of the social and cultural power of natural science as a model for all compartments of human knowledge. Nevertheless, so these scholars argue, it is the wrong model. In fact, some used the expression "scientism" to label both the exaggerated respect for natural science found among quantitative social scientists, along with the stereotypical treatment of it to be found in one particular orthodox picture of natural science, that of empiricism.

There are, according to these critics and other social commentators, other ways of knowing besides the methods that natural science employs. These critics defend as epistemically respectable the disciplines that others have stigmatized as pseudoscience—such as astrology, parapsychology, the theories that stand behind alternative "holistic" therapies in medicine, like homeopathy; and non-standard cultivation practices—such as playing music to one's house-plants. On their view to deny these paradigms epistemic status is simply to argue from the blinkered and question-begging perspective of the Newtonian paradigm, a paradigm for that matter now superseded by scientific advances in cosmology and quantum physics for which we have as yet no acceptable philosophical interpretation. Who can say whether or not, when the dust settles in these areas, alternative non-Newtonian ways of knowing will be vindicated?

To the extent that the visibility of the social study of science deriving from Kuhn has undermined the credentials of traditional natural science, it made more controversial the public support for the sciences in those countries, especially Great Britain in the 1980s, where the "strong program" in the sociology of science was most visible and intellectually influential.

Postmodernism and the Science Wars

There are further critics of scientism beyond the historians and sociologists of science and the writers of "new age" trade-books. Even scholars in the humanities, professors of English, French, and kindred disciplines, have sought to "de-center" science, and to treat its products as "texts" in the way that such scholars would treat Dickens' *Great Expectations* or Flaubert's *Madame Bovary*. The reason they offer for equivalent treatment of scientific and literary works, including those labeled by their authors as "fiction," is of course that in the end the difference between works purporting to describe the world and those with other aims is purely a social construction. These scholars often describe

themselves as "postmodern," a name to be contrasted with "modernism"—the now obsolete, outmoded, and discredited last gasp of a tradition that stems from the scientific revolution of the seventeenth century, which was continued through the Enlightenment of the eighteenth century and the romanticism and nationalism of the nineteenth century, and resulted in the horrors and consequent disillusionment of the twentieth century. Many of these postmodernists describe their method of work as "deconstruction," which reflects their dual aims of showing, first, that claims which purport to be grounded on and reflect reality are actually social constructions, and, second, that these claims should be suspect owing to the ways in which they conveniently support, favor, enhance, and reinforce the social, political, economic, racial, gender, or other interests of their advocates.

The tools that postmodernists equipped themselves with are largely those fashionable in Paris in the last quarter of the twentieth century, and are associated with names such as Derrida, Althusser, Lyotard, and, to a lesser extent, Foucault. Expounding these theories is a task beyond the powers of the present authors, but their implications are often recognizable extensions of themes in the work of Feyerabend, and can even be understood as conclusions suggested in the study of Quine and Kuhn. Of course neither Quine nor Kuhn would accept these conclusions as validly inferred from their doctrines, but both are safely dead.

Kuhn undercut the possibility of objective foundations for knowledge in observation, and Quine rejected any other source of certainty, especially as provided by fixed linguistic meaning. The French postmodernists and their allies took such doctrines, especially the linguistic ones, much further. According to them, underdetermination of theory by observation extends from physics to everyday life and of course to the meaning of our language. Anything anyone says is underdetermined, not least by the speaker's own meanings, as there are no such things as meanings—either thoughts in the head, or socially fixed meanings outside of people's thoughts. There is, in fact, no fact of the matter about what anything means. Accordingly it will be impossible unambiguously to identify the components of Kuhn's incommensurable paradigms, not just because there is no paradigm-neutral location from which to do so, but because there is also no authority within any paradigm about its meaning. There are of course competing claims about the meaning of a paradigm, indeed about the meaning and significance of any body of belief. But none is correct, and which among them secures local "hegemony" is a matter of social, political, economic, or other sorts of power.

Postmodernists often prefer the notion of "narrative" over paradigm, since its meaning is apparently fixed enough in scholarly discourse to suggest that general laws, theories, methodologies, philosophies, and all other discursively expressed objects of thought, are really in the end "stories" we tell in order to convince or amuse one another in the "conversations" that constitute each discipline.

The traditional view of science of course favors a "totalizing" narrative, one in which either the whole truth about reality is ultimately to be given, or in which the complete tool kit for predicting our future experiences can be constructed.

Both of these versions of the totalizing narrative seek to subsume all stories (the "total" narrative) by employing words like "universality," "objectivity," "essence," "unity," along with "truth" and "reality." Of course these expressions are merely sticks with which to beat into submission those who dissent from scientists' (and their philosophical fellow-travelers') orthodoxy. Once we recognize that these inscriptions and noises ("the truth, the whole truth, and nothing but the truth") have no fixed meanings, the claim that science employs them are open to doubt. It is only by wresting the power to influence the audience from the totalizing narrative of science that it can be replaced by other narratives, that will emancipate those social groups whose interests are not served by science or at least science as it has hitherto been practiced.

Postmodernism's analysis is of course not limited to science, and its tools are equally applicable to other social institutions—formal and informal—that fail to reflect the radical differences and incommensurable discontinuities among and between people. These differences do not require reconciliation into logically consistent packages: there is no transcendental logic on which to ground consistency, and in any case consistency is just a part of the totalizing narrative of science that we need to surrender! Contradiction is to be expected and self-contradiction is at most an unintentional, or for that matter a perfectly intentional, source of amusement and irony. Postmodernism is, however, consistent enough to insist that the excluded social groups, which the totalizing narratives render invisible, would and will immediately marginalize other groups, when they find themselves in positions to advance their narratives. The key thing to remember is that there is no fact of the matter about competing narratives, their interpretation, or their meanings.

What is the most generous assessment that one can make of this "take" on science?

Consider the enduring and fundamental problems that the analytical philosophy of science has raised for itself about the truth of scientific theories and the reliability of scientific method, about the metaphysical foundations of the sciences. All of these tend to undermine simple and unqualified confidence in any particular scientific achievement. And the fallibility of science can't be eliminated. Taking the risk to be mistaken is essential to science's capacity to enhance our understanding. These difficulties might be supposed by some to expose science to a certain amount of "demythologizing" if not debunking.

Add to this the undeniable fact that like any other human institution, science as it is practiced reflects the impact of individual and group interests. This must result in imperfections and failures due to human error and venality, and the imposition of broader cultural, religious, economic, and political forces from which the institution of science cannot protect itself. To the extent that scientists seek to insulate their practice from forces that tend to deform and deflect it from its commitment to "objectivity" it is important that the forces be identified. To the extent that the "strong program" itself adopts the methods of controlled inquiry characteristic of science it stands the chance of accurately identifying these deforming forces. Thus, scientific institutions may to some extent be

reshaped to mitigate their impact. But it remains unclear whether the "strong program" can consistently identify its own findings about science as true or reliable, and therefore a basis for application.

As for the more obscure and extreme assessments of postmodern deconstruction, it is by no means clear that any of their "results" should be taken seriously. For they appear to be a combination of sheer physics-envy or science-envy and a sort of *jeu d'esprit*—a game played among the European literati to amuse themselves. Let us see why this diagnosis seems right.

Does the Sokal Hoax Prove Anything?

If empirical scientists are still reading at this point, they may well be excused for not taking the contemporary humanities' approach to science very seriously. In fact, if they know much of the social history of postmodernism's encounter with real science, they may have excellent reason to treat its deconstruction of modern science as an empty game. These reasons were given to them by a physicist, Alan Sokal. Like others, Sokal recognized in postmodernism's position a similarity to that of the emperor in Hans Christian Andersen's "The Emperor's New Clothes." In that story the emperor walked about naked, and no one drew attention to this fact because it served their interests not to do so. Postmodernism has certainly been on the correct side of the "barricades" in modern intellectual life, opposing inequalities of all sorts, including racialism, social class exploitation, sexism, homophobia, undercutting stereotypes, expanding the range of artistic, behavioral, social, political possibilities people can visualize. And to the degree that the Newtonian or Darwinian or some other tradition in science has been employed to foster such inequalities and blinker such visions, as well as diminishing the importance of their cultural contributions, humanists sought tools to fight back. Having pretty well surrendered all of their proprietary literary and aesthetic theories, as well as their canon owing to its hegemonic, racist insensitivity to non-Western culture, they were particularly susceptible to a fashionable French doctrine that could enable them to "diss" the sciences. The patent unintelligibility of this theory was no obstacle, of course, for its technical apparatus, neologisms, jargon, and special symbols could function to protect it from the uninitiated just as mathematics functions for natural science.

In the 1990s increasing attention to work of this type, by journalists and social commentators, produced "the science wars"—an overdramatization for a debate in which one side (the humanists) made outrageous claims and the other side (the scientists) pretty much ignored them. The few who felt the need to defend science's objectivity and integrity were eager to show that the attack on it was "politically motivated," appallingly ignorant of science, and not to be taken any more seriously than a *jeu d'esprit*, an intellectual game played by humanists because they had nothing left to critique in the humanities.

Enter Alan Sokal. In 1993 the avowedly postmodern academic journal *Social Text* announced that it would publish a special issue on science. To test the hypothesis that the postmodern attack on science was carried on

by scientifically illiterate persons who were not intellectually serious, Sokal responded to this announcement by preparing and submitting a phony paper, an intentionally exaggerated satirical caricature of the sort of scholarly paper that deconstructionists write, entitled "Transgressing the Boundaries: Towards a Transformative Hermeneutics of Quantum Gravity," which made a number of outrageous claims about physical reality. The paper's intentionally invalid and unsound argument employed accurate quotations from the works of important postmodernist theorists, and concluded that the character of contemporary theory in quantum gravity (one of the most difficult and unsettled areas of physics) substantiated a set of aesthetic, ethical, and political values congenial to postmodernism.

The paper was accepted and published by *Social Text*, after which Sokal admitted his deception. How could this have happened? It turned out that the journal did not subject submissions to refereeing. The editors defended this unusual practice as one that fostered creativity and complained that they had carefully read the paper and requested revisions that the author had refused to undertake. The editors said that they had decided to publish the paper without the requested revisions to encourage the author. They complained that *Social Text* had been the victim of a "scam" and that Sokal's deception had violated the canons of academic honesty. The hoax was subject to scrutiny and discussion among sociologists and other students of science in books and conferences for several years thereafter. Few philosophers of science took notice of the matter in print.

Chapter 11 revealed the problems surrounding the notion of a "positive instance" of a hypothesis. These problems add to the difficulty of saying that the Sokal hoax provided a positive instance that to some extent confirmed the hypothesis of political bias and want of seriousness among postmodern critics of science. But what is clear is that it had a significant impact on at least several of the most prominent sociologists of science, especially the proponents of the "strong program." In the period from the late 1990s onward, several of the leading figures of this movement disassociated themselves from the social constructivist critique of science. Some even began to fear that their claims played into the hands of governments attempting to reduce the influence of science in social policy making (especially in the United Kingdom) and the hands of political and religious forces seeking to supplant scientific information with religiously motivated mis- and disinformation (especially in the United States). As a result, by the turn of the twenty-first century the science wars came to an amicable conclusion with philosophers of science and most social scientists studying scientific institutions and activities agreeing on its claims to epistemic reliability.

Recall the claim in Chapter 1 that the philosophy of science takes priority over other academic disciplines when it comes to the questions of understanding and assessing claims to scientific knowledge and the reliability of those methods to secure it. The work of the sociologists of science examined here reflects the grounds of that claim. Insofar as they can claim to supersede scientists' or philosophers' interpretations of what science is all about, it will be because the

scientific theories they bring to bear on the understanding of science can be certified as reliable. But this is something that presupposes answers to questions about science that are advanced and answered by philosophers of science.

This leaves two important matters to be dealt with. First is the question, which can be separated from the deconstructionist mantra, of how, where, and to what extent science is deformed by powerful interests intent on their own hegemony; the second is the examination of what went wrong in the post-positivist period that led intelligent and well-meaning people to take seriously such grave doubts about the objectivity of science. The first question is addressed in the next section. The second is an issue that the philosophy of science is not equipped to address.

Scientism, Sexism, and Significant Truths

It doesn't take a postmodernist to notice that science and scientific findings have been long misused in two ways. First, science as an institution has persistently provided more efficient and effective ways of harming people, other organisms, and the environment. Second, it has done so in part by providing unwarranted rationalization for policies that effect such harms. These trends must be granted even among the "friends" of science, indeed even among those afflicted with scientism. The trends enjoin an obligation among scientists and others who may influence the future of science to reduce as much as possible these untoward consequences in the future.

Among the most influential students of science committed to the improvement of science as a social institution have been feminist philosophers of science. Some of these philosophers begin their examination of science from an epistemological insight, sometimes called "standpoint theory." This theory begins with the uncontroversial thesis that there are certain facts relevant to the assessment of scientific theories that are only detectable from certain points of view—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 racial minority, or having a certain sexual orientation. 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 the woman or the minority or the relevant social class, the male would not detect the same fact; rather it must claim that they cannot detect such a fact for the same reason they cannot be female. The fact must evidently be a relatively complex, perhaps historical, theoretical fact not open merely to one equipped with the five senses. And feminist standpoint theorists have not been reluctant to identify such facts.

Typically they are facts hard to quantify, or even fully to describe in ordinary or scientific vocabularies. Examples include facts about the long-term effects of oppression, subordination, discrimination, stereotyping. These are hard facts and undeniable ones, for all the difficulty there may be describing them, and they can lay claim to being facts inaccessible merely from description, or brief

and/or simulated personal encounter. One has to live the standpoint to really detect the relevant facts. It is plain that these claims are particularly relevant in the social sciences. Few standpoint theorists allege that physical or chemical facts are missed by failure to attend to the findings from a women'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 non-human species and the failure to notice female strategies (allow access to males with best genes and with a demonstrated willingness to commit resources to offspring) was owing to male biologists' incapability of locating themselves at the relevant standpoint.

This example of course reflects the philosophical difficulty facing standpoint theorists. Opponents of this theory will argue that all it took was for female sociobiologists to draw the attention of their male colleagues to the facts for the entire discipline to revise theory in order to accommodate the facts. What standpoint theorists need to do is very difficult: On the one hand they need to identify both the facts inaccessible from other standpoints in a way that forces those occupying the other standpoints to grant the facts' existence, and they need at the same time to argue that they 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 **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 ones congenial to the empiricist and/or naturalistic views of contemporary non-feminist philosophy of science. The aspirations of standpoint theory in question include those of emancipation, not just of women, but of all who have suffered from the failures of "objectivity" and "disinterestedness" that science officially may extol but scientists actually fall short of.

Many feminist philosophers of science have been heavily influenced by Quine and by (a naturalistic interpretation of) Kuhn. Thus, they are prepared to identify facts that male scientists have missed not as in principle inaccessible to them, as standpoint theorists allege, but still not much accessed by males. But naturalistic feminists recognize that such facts require substantial theory to recognize, which the non-scientific interests, values, or even tastes of scientists brought up in a sexist world have probably prevented them from seeing. On the view of these feminists, theories, research programs, paradigms, are not incommensurable, but they are often impervious to any but a very forceful counterevidence wielded in politically effective ways.

Perhaps because feminist philosophers 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, traditional philosophy of science has embraced a conception of science as the enterprise of individuals—Kepler, Galileo, Newton, Lavoisier, Darwin, Einstein. In this they

have perhaps been overly influenced by the Cartesian tradition in epistemology, one that begins with Descartes' solipsistic skepticism and 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, institutions and governments. Feminists have noted both the strengths and weaknesses of this fact about modern science. On the one hand, the scientific community often serves to distribute research tasks in efficient and coherent ways, to support and scrutinize findings and theories that individuals advance, and to provide a reward (and punishment) structure that gives scientists incentive 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 scientific inquiry, and of its gendered deformation. Feminist philosophers argue that doing so should have an impact on its future and our philosophical assessment of it.

Empiricists usually distinguish facts from values and observe that science has long been characterized by a commitment to "value-freedom." It is 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 may require that we can distinguish factual judgments from value judgments up to the standards that Quine, for example, set for real distinctions in philosophy: in particular non-circularity in drawing the fact/value distinction. Some philosophers, both feminists and non-feminists, believe this is impossible. Others, as we shall see, claim that in any case making value judgments in science is unavoidable so that the attempt to rid science of such claims is a mistake.

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 it 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. But, first of all, 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 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 hitherto, nor how it should proceed hereafter. 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 other philosophers, feminist philosophers of science recognize that theory is underdetermined by observation: The direction of scientific theorizing

over time is not driven by experiment and observation alone. All or most scientific beliefs are insulated from direct observational challenge by the network of other statements, assumptions, and auxiliary hypotheses a scientist believes. Quine assumed that auxiliary hypotheses would be factual claims. But, following Lynn Nelson's *Who Knows: From Quine to a Feminist Empiricism*, some feminist philosophers have argued that, along with other factual assumptions, auxiliary assumptions include value judgments that also play a role in fixing beliefs otherwise underdetermined by evidence. If we cannot draw a distinction between factual claims and value judgments, this claim will be in little need of defense. Even if we can, there seems an attractive argument for the claim that values are inextricably bound up in science.

Like all intentional human activities, scientific activity is determined not just by what we believe, but also 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. Consequently we will never make so much as a dent in the number 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 so 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 dominated science; similarly, many lines of inquiry are absent from its history because of these same values. It is easy to give concrete examples of 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 infra-humans, 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 male samples 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 vast blind spots and blank spaces, which have resulted from 2,500 years of male domination in the identification of what questions are significant and which are not. What science needs to do now, or rather what women have always needed science to do, is to treat research questions that are significant to women. And the same goes for any other group, class, race that has been disposed of in the identification of significant and insignificant research questions.

The crucial point in this argument is not that science should forgo judgments of significance. It cannot do so. There are too many research questions to choose from in science's search for truth. 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 science is not the end of the feminist agenda in the philosophy of science. It is probably closer to the beginning of it. Feminists have argued further that the besetting sin of scientism is that of mistaking masculine styles of scientific inquiry for all scientific inquiry. Thus they have argued, for example, that demands for unification in scientific theorizing and explanation are often premature, counterproductive of scientific progress, or unreasonable even in a mature discipline. Feminist philosophy of science encourages "pluralism." Women, and the science they pursue, are more prepared than traditional male-dominated science to tolerate multiple, competing, complementary, and partial explanations, without the expectation of near-term weighting of importance, placement in a (patriarchal) hierarchy of causes, or unification under a single complete theory. This ability to tolerate and willingness to encourage a variety of approaches to the same scientific problem reflects a 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 science itself, the feminist commitment to pluralism should be equally embraced by all, at the evident expense of the totalizing and reductionistic proclivities of more traditional science. 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 cannot after all be a matter of complete disinterestedness, or 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.

Similarly, some feminist philosophers of science reject the centrality of prediction, and especially of control, to the scientific enterprise. The suggestion that science optimally should proceed in this way reflects what they hold to be masculine biases that are also reflected in the subordination of women and other marginalized groups. The methodology of prediction and control fails to gain knowledge that might derive from a more cooperative relationship with the objects of scientific study, be they human or infra-human. 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 an innocent one. 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 independently 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 science are the empiricists and naturalists among them. 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 non-feminist philosophy of science. Unlike postmodernists and others who take up an adversarial stance against scientism, these empiricist feminists do not challenge science's aim to provide objective

knowledge, but seek to broaden our understanding of what objectivity consists in and how more nearly to attain the goal of objective knowledge. Accordingly these philosophers and those who share their agenda still need to come to grips with the arguments of those who have embraced the more radical epistemic relativism that has characterized much of the post-Kuhnian study of science.

Summary

Sociologists and others eager to reduce the baleful influence of a blinkered, narrow-minded, patriarchal, capitalist, and probably racialist paradigm associated especially with Newtonian science have adopted Kuhn's view of science as a version of epistemological relativism.

Relativism in epistemology, as in ethics, allows for the possibility of alternative and conflicting views without adjudicating which is objectively correct: none are, or rather each is correct from the perspective of some epistemic point of view, and all points of view have equal standing. So far as the strongest sociological interpretation of Kuhn was concerned, science is moved by social forces, not epistemic considerations. Science is a social institution, like any other; and this is how it is to be approached if we wish to understand it.

If the empiricist criticizes this argument as incoherent, the relativist is indifferent. All the relativist requires is an argument that convinces relativists, not one that is even intelligible to, let alone accepted by, the empiricist. But this is the end of all debate, and in recent years many of the most radical of sociologists of science have given up this degree of relativism.

And of course many philosophers of science, especially feminists among them, have sought to gain from some of the social studies of science an improved understanding of how it proceeds and how it may more effectively secure its objectives, while avoiding the relativist's conclusions.

Study Questions

1. "Poetry is untranslatable. Science is not. Therefore, incommensurability is false." Sketch an argument for this view.
2. "Science must be open-minded. It should welcome methodological anarchism." True or false?
3. Defend or criticize: "Philosophy of science is part of the problem in really understanding what science is, not part of the solution."
4. Could humanists have pulled off a "Sokal hoax" on scientists?
5. Can the feminist critique of male-dominated science really be reconciled with its claims to disinterestedness and objectivity?
6. Can we give an account of science as a search for significant truths that ensures its freedom from the deforming effects of bias, partiality, and special interests?

Suggested Readings

The classical text predating Kuhn's influence in the sociology of science is R. K. Merton, *The Sociology of Science*. Steven Shapin, *The Scientific Revolution*, is a good introduction to the history of the critical period of the seventeenth century. See also the suggestions at the end of Chapter 13. Sokal reported on his hoax and the controversy it stirred in A. Sokal and J. Bricmont, *Intellectual Impostures*.

Many of the works, especially collections of papers, about Kuhn's books mentioned in the last chapter are of great relevance here. Among the most radical of relativist sociologists of science in the period after 1970 are B. Latour and S. Woolgar, *Laboratory Life*; A. Pickering, *Constructing Quarks*; Shapin and Schaffer, *Leviathan and the Air Pump*; B. Barnes, *Scientific Knowledge and Social Theory*; and D. Bloor, *Knowledge and Social Imagery*. Bloor and Barnes significantly qualified their views 20 years later in B. Barnes, D. Bloor, and J. Henry, *Scientific Knowledge: A Sociological Analysis*.

Unsympathetic accounts of relativist doctrines about science and their impact include N. Koertge (ed.), *A House Built on Sand*, and Gross and Levitt, *Higher Superstition*. Readers may consult these two works to identify sources advocating the views that these authors attack.

Among important works in feminist philosophy of science is S. Harding, *The Science Question in Feminism*. Harding and O'Barr, *Sex and Scientific Inquiry*, anthologizes important contributions by feminist philosophers of science. *Hypatia*, vol. 10, 1995, includes several papers by feminists in the empiricist tradition. One of these papers, E. Anderson's "Feminist Epistemology: An Interpretation and Defense," is reprinted in Balashov and Rosenberg. Another work in this tradition is L. Nelson, *Who Knows: From Quine to a Feminist Epistemology*. Important work in the philosophy of science sympathetic to both the sociological approach and the defense of objectivity is due to H. Longino, *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. *Scrutinizing Feminist Epistemology*, edited by Pinnick, Koertge, and Almeder, provides an anthology of recent papers, several of which are unsympathetic to feminist epistemology. Harding has edited *The Feminist Standpoint Reader*, which includes much more supportive essays on feminist epistemology. Much the best work in the field is anthologized by L. Anthony and C. Witt in *A Mind of One's Own*.

Curd and Cover reprint Longino, "Values and Objectivity," and Kathleen Okrulik, "Gender and Science."

15 Science, Relativism, and Objectivity

Overview

Throughout its history, the philosophy of science has tried mightily to find a way to justify the special epistemic status of science. Many challenges have been considered and addressed, but perhaps the most far-reaching of these is the relativist's challenge to science as a distinctive body of knowledge, which attains higher standards of objectivity and reliability than other methods. Dealing responsibly with this notion requires that we return to the fundamental problems in epistemology, the philosophy of language, and metaphysics, in order to see where philosophy might have gone wrong and what led the radical followers of Kuhn to conclusions of such patent preposterousness. It may also require that we attend to the findings of relevant sciences, such as cognitive and perceptual psychology, to discover whether there are theory-free sources of data and hypothesis formation in our psychological make-up.

Relativism and Conceptual Schemes

For all of Kuhn's insights into the history of science, most philosophers of science consider that something has gone seriously wrong in the development of the social studies of science that grew up after his work. Much of the motivation for the attempt to understand natural science stems from a (perhaps sexist) appreciation of its predictive power and explanatory depth. A related motivation

stems from the (arguably "totalizing") desire to identify its methodological tools so that they can be applied elsewhere (especially in the social and behavioral sciences) with the same theoretical insights and technological results. When an inquiry so motivated concludes that science is just another religion, just one of a wide variety of ways of looking at the world, none of which can claim greater objectivity than the others, then sometime, somewhere we have taken a wrong turn.

But where? It is simply not enough to turn one's back on Kuhn's insights, nor on the arguments against the pretensions of science mounted on top of them. Many philosophers of science have concluded that Kuhn's historical account of scientific change has been "over-interpreted"; that he did not intend *The Structure of Scientific Revolutions* as a broadside attack on the objectivity of science. In this they had the support of Kuhn, at least while he lived. On numerous occasions, Kuhn said that it had not been his intention to cast science down from its claims to objectivity, but to enhance our understanding of it as a human enterprise. Similarly, Quine and his philosophical followers could not countenance the misapplication of their doctrine of underdetermination to support the conclusion that current scientific conclusions are not the most reasonable and well-supported conclusions we can draw about the world. But what Kuhn and Quine may have intended cannot decide what their arguments have in fact established or suggested.

What the defender of scientific objectivity, or at least its possibility, must do is undermine the claims of incommensurability. To do this one must either attack the assimilation of observation to theorizing, or reconcile it with the possibility of testing theories by observation in a non-question-begging manner. And to show how science can make progress over theoretical change that cumulates knowledge, we will have to show how translation between theories can be effected.

One way that defenders of objectivity in science have attempted to reconcile the assimilation of observation to theory with its continued role in testing is to draw a distinction between the categories we adopt for classifying particular items—objects, processes, events, phenomena, data—and the particular acts of classification themselves. Differing and even incommensurable categorical frameworks can be reconciled with agreement about actual findings, thereby making objectivity in the recording of data possible. The difference is like that between the letter-box pigeon holes in a departmental office and the particular pieces of mail that are distributed to them. Adopting a particular set of labels for boxes doesn't prejudge what pieces of mail will come in. Observations are like pieces of mail. Their descriptions are the labels on the classes into which we sort observations. A hypothesis is a claim that members of one category will also fit into another, or always come together with members of another category. There may be agreement on what falls into any category, and thus a way of testing hypotheses, even when the hypotheses are expressed in terms of categories controlled by a theory that is not itself tested by what falls into them. It can even turn out that differing categorical schemes will substantially overlap, thus allowing for agreement about data even between differing categorical frameworks. For

example, items that the categorical framework of Einstein's theory of special relativity would classify as "having mass" would also be so classified by Newton's theory, notwithstanding the fact that the two theories mean something quite different by "having mass." And of course, we may surrender categorical systems when they no longer work well, that is, when it becomes difficult to use them to file things uniquely, or too complicated to figure out in which boxes they belong. Thus, observation can control theory even when its most basic descriptions reflect pre-established theories, even theories we don't recognize as theories, like those embodied in common sense and ordinary language.

But when one thinks about the notion of a categorical scheme and instances that are classified in accordance with it, the conclusion that there is a place for theory-controlling observations here is simply question-begging. To begin with, items don't come with labels that match up with the labels on the categories: samples of gold don't have the word "gold" printed on them. The simplest act of classification requires hypotheses about other categories. Classifying something as gold requires that we invoke the hypothesis that gold dissolves only in aqua regia. This hypothesis presupposes another set of hypotheses that enable us to tell what aqua regia is. And so on, ad infinitum. The "ad infinitum" is due to the fact that there is no basement level of words defined directly by experiences, as the historical empiricists held.

Second, how do we tell the difference between hypotheses about correlations between items in our classifications, like "gold is a conductor" and hypotheses like the one about gold and aqua regia that we need to do the classifying? We need to be able to tell the difference between these hypotheses if we are to treat one set as open to objective test, while the other is not, owing merely to its classificatory role. We can't argue that classificatory statements are true by definition (gold = whatever dissolves only in aqua regia), and that the "gold is a conductor" hypothesis is a claim about the world. Without first having established a way of empirically telling the difference between definitions and factual claims we simply cannot do this, and doing this would require still another argument against Quine.

Third, categorical schemes are in fact hypotheses about the world, so the whole distinction breaks down. Consider the most successful categorical scheme that science has ever established: Mendeleyev's Periodic Table of the Elements. It is a successful categorical scheme because it purports to "divide nature at the joints." The differences between the elements it systematizes are given by atomic theory. In the century after Mendeleyev advanced his categorical system, discoveries especially about nuclear structure and electron-shell-filling explained the relationship between Mendeleyev's rows and columns, and showed that it was more than a merely convenient filing system. It was a set of hypotheses about similarities and differences among elements—known and unknown—that required further and deeper explanation.

Fourth, and finally, it is pretty clear, especially in the case of fundamental theories or paradigms, that any disagreements are not about the individual instances and which categories they are to be filed in. Rather, the disagreements are about

the definitions of the categories that make these agreements about classifying impossible, and cannot be compromised: compare Aristotle and Newton on what counts as "rest." Differences in classification reflect incommensurabilities that preclude theory-comparison.

Acceding to the assimilation of observation to theory, while distinguishing categories from their instances, will not preserve the objectivity of science. Rather, the defender of scientific objectivity will have to seek out countervailing evidence from the history of science, and better psychological theory and data that counter the psychological claims on which the denial of the distinction between observation and theory rests. Such evidence might show that all humans have some common inherited sensory categorical scheme shaped by evolution to be adapted to success at science or some other enterprise that science can make use of. This is certainly one approach that has been adopted, especially by naturalists. It is open to the question-begging objection of course: Appealing to findings and theories in psychology is itself to adopt a non-observational and therefore non-objective basis from which to criticize opposition to objectivity. But then, this is the same kind of evidence that Kuhn and his followers originally cited to undermine the observational–theoretical distinction.

Such opponents of objectivity cannot have it both ways. Indeed, one might even charge them with the deepest form of incoherence, for they purport to offer arguments against the objectivity of science. Why should we believe these arguments? Do they constitute an objective basis for their conclusions? What makes their arguments and evidence probative, when the arguments of their opponents are always question-begging? These rhetorical questions do not carry the debate very far. This is largely because opponents of scientific objectivity have little interest in convincing others that their view is correct. Their dialectic position is largely defensive; their aim is to protect areas of intellectual life from the hegemony of natural science. To do so, they need only challenge its pretensions to exclusivity as a "way of knowing." These opponents of scientific objectivity cannot and need not argue for a thesis stronger than epistemic relativism.

The strongest card of the opponent of scientific objectivity, therefore, is the incommensurability of meanings that insulates paradigms and theories even from inter-translation. Incommensurability means that no critique of any theory from the perspective of another is even intelligible. Again, it is not enough to call this doctrine self-refuting, on the ground that in order to communicate it to someone with whom prior agreement has not been established the doctrine must be false. Such a *reductio ad absurdum* argument is a matter of indifference to opponents of objectivity in science, who are not interested in convincing others but in defending their own view as invincible.

Dealing with Incommensurability

One apparently attractive alternative to the *reductio* argument begins by drawing attention to a fundamental distinction in the philosophy of language between meaning and reference. Meanings, all will admit, are a great difficulty

for philosophy, psychology, and linguistics; but reference, or denotation, or the extension of a term, seems less problematic. What a word names, what it refers to, is something out there in the world, by contrast with what it means, which may be in the head of a speaker and/or a listener, or a social rule or convention, or a matter of use, or as Quine and his followers might have it, nothing at all. And because the reference of a term is something out there, as opposed to in here (pointing to the head), speakers may agree on what a term names without agreeing on what the term means. In the case of terms that name properties instead of things, like "red" or "loud," we can agree on the instances of things and events that bear these properties. The things that are instances of "red" or "sweet" or "rigid" are members of the "extension" of the term "red" or "sweet" or "rigid." We can agree by inspection on whether things are in the extension of "red" or not, even when we can't get into one another person's head to find out whether what looks red to you looks red to me. We can agree that "Superman" names the same item as "Clark Kent" without concurring that the two expressions have the same meaning (indeed, proper names, like "Clark Kent," have no meaning). Reference and extension, it may be held, are more basic and indispensable to language than is meaning. Moreover, it is tempting to argue, in the manner of the empiricists of the eighteenth century, that language cannot be learned unless it starts with terms that have only reference or extension or something like it. For if every term has meaning—given by other words—it will be impossible for a child to break into the circle of meaningful terms. To break into language, some words must come to us as understandable solely by learning what they refer to, or at least what events stimulate others to use them.

Finally, there are good arguments to suggest that what is really indispensable for science and mathematics is not that the meanings of terms be given, but that their references be fixed. Take any truth of arithmetic, for example, and substitute any term within that preserves its reference, and the statement will remain true. For example: $3^2 = 9$ remains true when it is expressed as the square of the number of ships in Columbus' 1492 fleet. If two scientists can agree on the reference of terms, or on the set of things a scientific term is true of—for example, the set of things having mass, whether Einsteinian or Newtonian—they need not agree on the meaning of the term, or whether a translation is available from one meaning of the term to another. Could agreement on reference be enough to ensure commensurability between scientific hypotheses, theories, or paradigms? So some defenders of objectivity, following Israel Scheffler, have argued.

Suppose inquirers could agree on the reference or extension of a set of terms, "F" and "G," without even discussing their meanings. Suppose further that this agreement led them to agree on when the extensions of these terms overlap, or indeed are identical. In the latter case, they would have agreed that all Fs are Gs, without even knowing the meanings of "F" or "G." Such meaning-free agreement could be the basis for comparing the differing theories that inquirers may embrace, even when these theories are incommensurable. A set of hypotheses about the correlation among objects named by categories on whose reference

scientists agree would provide exactly the sort of theory-free court of final authority that would enable us to compare competing and incommensurable theories. Each hypothesis on which scientists concur under their purely referential construal would be given different meanings by one or another incommensurable theory. But it would be an objective matter of mathematical or logical fact whether, thus interpreted, the hypotheses would be derivable from the theories to be compared. The theory would be best supported that deductively implied those hypotheses on the extension of whose terms there was agreement.

It doesn't take much thought to realize that the only hypotheses that will qualify as purely referential will be ones about objects on which agreement of reference can be established non-linguistically, i.e. by pointing or otherwise picking out things and properties without words. But the only candidates for such hypotheses will be those expressed in the vocabulary of everyday observations! In other words, the appeal to reference is but a covert way of bringing back into play the distinction between observational and theoretical vocabulary that started our problem. One way to see this is to consider how we establish the reference of a term. Suppose you wish to draw the attention of a non-English-speaker to an object on your desk, say an apple. You could say "apple" but to a non-English speaker that will not discriminate the apple from anything else on your desk. Suppose you say "that" or "this," while pointing or touching the apple. Well, that will probably work, but it is because your interlocutor knows what an apple is and has a word for it. Now, suppose you wish to draw your interlocutor's attention to the stem of the apple, or the soft brown spot under the stem, or the worm wriggling out of the soft spot, or the depression just under the stem. How might you go about it? What you do now is just about what you did the first time: you point and say the words. And that reveals the problem of working with reference alone. There is no way to tell what you are referring to when you say "this" and point. It could be the apple, the soft spot, the darkest part of the soft spot, the stem, the space occupied by the apple, or any of a large number of other things in the general vicinity of your index finger. Of course this is not a problem when we have other descriptive terms to individuate the particular thing to which we are in fact referring. But the reason this works is of course that these other words have meaning and we know what their meanings are! In short, without a background of meanings already agreed to, reference doesn't work. Pure reference is a will-o'-the-wisp. And the guide to reference is in fact meaning. The only purely referential terms in any language are the demonstrative pronouns—"this," "that"—and these fail to secure unique reference. Elsewhere in language the relation between reference and meaning is exactly the opposite of what we need. Securing reference relies on meaning. This is particularly apparent for scientific vocabulary, which is used to refer to unobservable things, processes, and events, and their only indirectly detectable properties.

If meaning is our only guide to reference, and the meanings of each of the terms of a theory are given by the role that the terms play in the theory, then theoretical holism about meaning makes reference part of the problem for the defender of scientific objectivity, not part of the solution. If theories or

paradigms come complete with categorical systems into which particular objects are classified, then proponents of two different paradigms or theories will not be able to agree on how particular things are classified, except by the light of their respective theories as a whole. This makes each of the theories recalcitrant to any experimental evidence that might disconfirm them. For in classifying events, things, or processes, the entire theory is involved, and the description of a counterexample to the theory would simply be self-contradictory. Imagine, given the meaning of the word "rest" in Aristotle's physics, the idea that an object could be moving in a straight line at constant non-zero velocity and have no forces acting upon it. Movement for Aristotle is *ipso facto* not rest, and requires a continually acting force. Nothing would count as being free from the influence of forces that was moving at all. Similarly, whatever it is that an Einsteinian might treat as disconfirming Newton's principle of the conservation of mass, it cannot be anything that a Newtonian could even treat as having mass.

But suppose that there were a way adequately to draw the distinction between observation and theorizing that could help us at least in principle to establish the possibility of translating across scientific theories and paradigms. Doing this would put us in a position to take seriously the problem of underdetermination. For the underdetermination of theory by data in fact presupposes both the observational/theoretical distinction and the comparability of competing theories. Quine certainly did not claim the universality of underdetermination in order to undermine the objectivity of science, only our complacency about what its objectivity consists in. But historians, sociologists, and radical interpreters of Kuhn's theory certainly have claimed that underdetermination means that, in science, theory choice is either not rational, or rational only relative to some social, psychological, political, or other perspective.

Defenders of the objectivity of science need to show that scientific changes are in fact rational, and not just relative to a point of view. They need to show that the changes in a theory that new data provoke are not just arbitrary, that the acceptance of a new paradigm is not simply a conversion experience, but is justified even by the lights of the superseded paradigm. To do this the philosopher of science must perforce become a historian of science. The philosopher must scrutinize the historical record with at least the care of a Kuhn, to show that beneath the appearances of "madness" that Kuhn and his successor historians catalogued there is a reality of "method." That is, philosophers need to extract from the historical record the principles of reasoning, inference, and argument that participants in paradigm shifts and theoretical change actually employed, and then consider whether these principles can be vindicated as objectivity-preserving ones. This is a task that naturalistic philosophers in particular have set for themselves. They have begun to wrestle with the archives, lab notebooks, correspondence, and papers of the scientists engaged in scientific revolutions great and small, and at the same time kept an eye on what the sciences, especially cognitive science, can tell us about reasoning processes that are characteristic of humans and the adaptive significance of reasoning for our ability to survive and thrive. As noted above, however, naturalists must at the same time take seriously

the charge of begging the question that dogs the attempt to preserve objectivity in the face of the holism of meanings and the want of a clear observational–theoretical distinction.

The charge of question-begging is central to the claims made by opponents of scientific objectivity and progress. They would hold that attempts to underwrite the traditional claims of science are not just paradigm-bound, but can be undermined by the very philosophical standards of argument and substantive philosophical doctrines that defenders of objectivity embrace. If correct, this situation provides a major challenge to those who seek both to understand the nature of science and to vindicate its traditional claims. The challenge is nothing less than that which faces philosophy as a whole: to articulate and defend an adequate epistemology, and philosophy of language. The task is then to show that episodes in the history of the sciences sustain these accounts of what constitutes knowledge and how reference can be secured to the same objects in the world by scientists with profoundly different beliefs about the world. If the philosophy of science has learned one lesson from Thomas Kuhn, it is that it cannot let the analysis of what actually happened in science fall exclusively into the hands of those with a relativistic or skeptical agenda.

Some scientists and proponents of "scientism" will be tempted to turn their back on these issues. They may well suppose that, if people who can't or won't do the hard work to understand science wish to pretend it isn't the best approximation to the truth about the world, this is their problem. And if there are people whose wish that there be a reality—religious, spiritual, holistic, or metaphysical—that transcends anything that science can know about, which leads them to the thought that science is blinkered and partial in its account of the truth, well, who are we scientists to wake them from their dogmatic slumbers? But the stakes for science and for civilization are too high simply to treat those who deny its objectivity in the way we would treat those who claim the Earth is flat.

Conclusion: The Very Idea of a Conceptual Scheme

As is evident from a survey of obvious moves in the attempt to restore the fortunes of an empiricist theory of knowledge and metaphysics, as well as an empiricist account of language, easy solutions will not avail. There is still much work to be done by philosophy if we are to understand fully the nature of science. Our project must include an understanding of categorization and observation, both philosophically and psychologically. We must clarify the relations between meaning and reference. We need an epistemology that adequately deals with the problem of underdetermination or shows that it does not obtain. And the philosophy of science must come more fully to grips with the history of science. These are all tasks for a naturalistic philosophy.

But perhaps we can take some comfort in a general argument offered by one of the most important and influential of Quine's students, Donald Davidson, that matters are not so desperate as they seem for a commitment to rationality that may seem embattled by relativism. In a famous paper, "On the Very Idea of

a Conceptual Scheme," Davidson offered a powerful argument against wholesale (or even very much) incommensurability. His aim was to prove that there could be but one conceptual scheme, or at least that we cannot distinguish multiple untranslatable linguistic and logical schemes from their content as it might be given by the world. This denial of a form–content distinction will be familiar as one of Quine's two dogmas. But Davidson puts it to use to show that translation between theories or paradigms must always be possible, because there is in reality only one conceptual scheme that is actual.

Davidson begins by noting that any sort of translation or interpretation must operate on a number of unquestionable assumptions. If the person before one is speaking a language, as opposed to merely producing noises, we have already to presuppose that most of the person's "common-sense" beliefs about non-controversial matters are the same as our own. If our translation of the noises and inscriptions the person produces results in the attribution to the speaker of nonsensical beliefs such as "furniture is liquid" or "clouds are colored red by numbers" we can be sure that our translation manuals are faulty. And of course it is not merely the beliefs, but also the many wants and desires of the speaker that must turn out to be the same as or similar to ours, once we match our translations of their sentences and match them up with the behavior they exhibit. When what the speaker says translates into our language as "arsenic is tasty and good for one's appetite" while the speaker is busy removing rat poison from a child's play room, again we can be confident that our translation manuals are at fault. The more one explores the constraints on adequate translation, the more it becomes apparent that no matter how different the structure and semantics of another person's language may be, if it is a language that they speak, then we have to make substantial assumptions about their rationality—and about the "metaphysics" or "ontology" of basic kinds of things, events, and processes than obtain in reality. In other words, the degree of incommensurability of their conceptual schemes with our own is at most quite limited. We can translate a great deal of anything anyone else believes, no matter how bizarre some of it may be, because most of it will not be bizarre. Fixed, agreed, common conceptual ground will enable us to pin down the area and nature of disagreement.

In this connection it is worth recalling the original context in which mathematicians introduced the term "incommensurability." It is a remainder that can be reduced to whatever dimensions we may consider negligible for our purposes. The assurance that Davidson's argument provides gives us confidence that even in the metaphorical extension of the mathematician's notion to the relationships between theories, paradigms, and world-views, the untranslatable remainder can always be reduced to negligible dimensions. Therefore, even if we are not yet able to solve all of the outstanding problems of epistemology and metaphysics, the threat of relativism may be reduced to levels that can be safely ignored.

Suggested Readings

A defense of classical empiricist theories of knowledge and language and of a realist metaphysics for science along the lines developed in this chapter is to be found in I. Sheffler, *Science and Subjectivity*. Nagel attacks Feyerabend's version of theoretical incommensurability in *Teleology Revisited*, as does Achinstein, *Concepts of Science*.

Davidson's argument can be found in *Inquiries into Truth and Interpretation*, a collection of his later essays.

H. Douglas, *Science, Policy, and the Value-Free Ideal*, and P. Kitcher, *Science in a Democratic Society*, explore the importance of objective scientific knowledge in the context of democratic decision-making.

There has been much interest by epistemologists in understanding relativism. J. MacFarlane, *Assessment Sensitivity: Relative Truth and its Applications*, is an advanced treatment of the issues.

Glossary

The introduction of each of these terms is highlighted in bold type in the main text.

A posteriori   Contrasted with *a priori,* meaning based on or justified by experience.

A priori   An *a priori* truth can be known without experience, i.e. its justification does not require knowledge about the way the world is arranged. For example, that 2 is an even number is a statement that can be known *a priori*. Note we may become acquainted with *a priori* truths through experience, but experience is not what justifies them.

Analytic truth   A statement true in virtue of the meanings of the word alone. For example, "All bachelors are unmarried males." Analytic statements can be known *a priori* (see ***a priori***). Philosophers following Quine are skeptical that we can distinguish analytic truths from some synthetic truths (see below) by any empirical or behavioral test.

Antirealism   The denial of scientific realism, according to which it is not reasonable to believe that the unobservable items in the ontology (see below) of any scientific theory actually exist, and that we should adopt an instrumentalist (see below) attitude towards theories which treat them as heuristic devices.

Aristotelian mechanics   The set of theories of motion advanced by Aristotle and his followers, including a commitment to purposive or teleological motion in the vicinity of the Earth and to all motion as requiring the imposition of forces creating impetus.

Axiomatic system   A set of axioms and their logical consequences, as proved by deductive logic. A statement is an axiom in an axiomatic system if it is assumed in the system and not proved. A statement is a theorem in the axiomatic system if it is proved in the system by logical deduction from the axioms. For example, Euclidean geometry begins with five axioms from which all the theorems are derived. The syntactic approach to theories (see below) holds that they are axiomatic systems.

Bayesianism   An interpretation of probability which holds that probabilities are degrees of belief, or betting odds, purely subjective states of scientists, and that probabilities are not properties of sequences of events in the world. Bayesians employ this conception of probability in order to explain and justify scientists' use of data to test hypotheses.

Boundary conditions   The description of particular facts which are required along with a law to explain a particular event, state or fact, according to the D-N model of explanation. Also known as "initial conditions." For example, in the explanation of the sinking

of the *Titanic*, the fact that the ship struck an iceberg of particular size at a particular velocity constitutes the boundary conditions.

Causation   The relation between events, states, processes in the universe which science sets out to uncover, which its explanations report and which its predictions about provide tests of its explanations. According to the empiricist analysis of causation, following Hume, the causal connection is contingent (see below) and consists in the instancing of regularities, and there is no connection of real necessity between cause and effect. It is widely held that causal sequences differ from accidental sequences, and that counterfactual conditionals (see below) reflect this fact.

***Ceteris paribus* clause**   From the Latin, "other things being equal." A qualification to a generalization that "if P then Q" which reflects the fact that other conditions besides P must obtain for Q to obtain. Thus, striking a match is followed by its lighting, but only *ceteris paribus,* for in addition to the striking, oxygen must be present, the match cannot be wet, no strong wind can be blowing, etc.

Constructive empiricism   The claim, due to van Fraassen, that theories are either true or false (realism) but that we cannot tell, and therefore should accept or reject them solely on the basis of their heuristic value in systematizing observations.

Contingent truth A statement whose truth is dependent on the way things actually are in nature, and not dependent only on purely logical or other grounds we could know about without experience. Contrast with necessary truth. Example: normal humans have 46 chromosomes (they could have had 48 or 44).

Counterexample   The identification of one or more items whose existence is incompatible with some statement and therefore a counterexample to its truth. Thus, a particle of finite mass traveling faster than the speed of light is a counterexample to the principle that nothing travels faster than light. One counterexample is sufficient to refute a generalization.

Counterfactual conditional   A statement of the grammatical form, "If P were the case, then Q would be the case," by contrast with an indicative conditional, "If P is the case, then Q is the case." When a counterfactual is true, even though the sentences contained in its antecedent and consequently (the P and Q) are false, then this suggests the two sentences P and Q report facts which are related as cause and effect, or are connected in a law.

Covering law model *See* **deductive-nomological model** of explanation.

Deductive-nomological (D-N) model   An explication of the concept of explanation which requires that every explanation be a deductive argument containing at least one law, and be empirically testable.

Deductively valid argument   An argument in which if the premises are true the conclusion must be true. For example: any argument of the form "If p then q, p, therefore q" is valid. The premises of an argument need not be true for the argument to be valid. For example, "All dogs are cats, all cats are bats, therefore all dogs are bats" is valid. Validity is important because it is truth-preserving: in a valid argument, if the premises are true (and of course they might not be), then the conclusion is guaranteed to be true.

Disposition   A trait of something which it exhibits only under certain conditions. Thus, glass has the disposition of being fragile, that is, it breaks when dropped from a certain

height to a surface of a certain hardness. Empiricists hold that dispositions obtain only when there are underlying properties that realize them. A glass is fragile even when it is never broken owing to the molecular structure of the material it is composed of. Dispositions without underlying structures that explain them are suspect to empiricists.

Empiricism   The epistemological thesis that all knowledge of non-analytic truths (see above) is justified by experience.

Epistemic relativism   The thesis that there are no propositions knowable, except relative to a point of view, and therefore no truths except relative to points of view. The epistemology associated with any one point of view has no grounds from another point of view.

Epistemology   The division of philosophy which examines the nature, extent, and justification of knowledge, also known as "theory of knowledge." The question whether we can have knowledge of unobservable things is an epistemological question. Compare **metaphysics**.

Exemplar   A term employed by Kuhn to characterize the standard textbook example of a solution to a puzzle dictated by normal science, or a particular piece of laboratory equipment along with the rules for its correct employment.

Explanandum (pl. explananda)   The statements that describe what is to be explained in an explanation.

Explanans (pl. explanantia)   The statements that an explanation of some fact consists in.

Explication (rational reconstruction)   The redefinition of a term from ordinary language which provides necessary and sufficient conditions in place of vague and imprecise meanings, and so eliminates ambiguity and the threat of meaninglessness. This method of philosophical analysis was advocated by the logical positivists. For example, the D-N model explicates the ordinary conception of "explanation."

Falsification   The demonstration that a statement is false by the discovery of a counterexample (see above). Popper held that the aim of science is to falsify hypotheses and to construct new ones to subject to falsification, since verifying scientific laws (see below) is impossible. If statements can only be tested by employing auxiliary hypotheses, strict falsification is impossible, for it is the set of auxiliary hypotheses and the hypothesis under test which is falsified, and not any one particular statement among them.

Feminist philosophy of science   The study of, and often the advocacy of, the thesis that gender does or should have an impact on the certification of claims as scientifically justified. Feminist theory sometimes suggests that some claims in science result from gender bias or neglect of gendered sources of formation, and as such are unreliable.

Function   The attribution to some trait, feature or component of a larger system of making a contribution to the attainment of some state of the containing system. E.g., the function of the bird's forward appendage is to fly. Functional attributions appear to be teleological: the bird has wings in order to fly.

Holism   The doctrine that scientific hypotheses do not meet experience for testing one at a time, but only in large sets, so that falsifications do not undermine one particular statement (see falsification) and confirmations do not uniquely support one particular set of statements (see underdetermination).

Hypothetico-deductivism   The thesis that science proceeds by hypothesizing general statements, deriving observational consequences from them, testing these consequences to indirectly confirm the hypotheses. When a hypothesis is disconfirmed because its predictions for observation are not borne out, the scientist seeks a revised or entirely new hypothesis.

Incommensurability   The supposed untranslatability of one theory or paradigm into another. If paradigms or theories are incommensurable, then there will be no possibility of reduction (see below) between them, and in moving from one to another, there will be explanatory losses as well as gains.

Inductive argument   An argument in which the premises support the conclusion without guaranteeing its truth, but contrast to a deductive argument. For instance, that the sun has risen many days in the past is good grounds to believe it will do so tomorrow, but does not make it logically certain that it will.

Inductive-statistical (I-S) model of explanation   An adaptation of the deductive-nomological model to accommodate explanations that employ probabilistic generalizations instead of strict laws. Probabilistic laws do not deductively entail the events they explain, and therefore the model differs sharply from the D-N model.

Inference to the best explanation   A form of argument employed in science to infer the existence of otherwise not directly observable or detectable mechanisms on the grounds that hypothesizing them best explains observations. A similar pattern of reasoning purports to establish scientific realism (see below) on the grounds that only the approximate truth of current scientific theories can explain the technological success of science.

Initial conditions *See* **boundary conditions**.

Instrumentalism   The thesis that scientific theories should be treated as heuristic devices, tools for organizing our experiences and making predictions about them, but that their claims about unobservable things, properties, processes and events should not be taken as literally true or false.

Logical empiricism   Synonym for logical positivism, which reflects the affinity of this school of philosophy to the British empiricists, Locke, Berkeley, and Hume.

Logical necessity   A statement is a logical necessity if its truth follows from the laws of logic alone, or if its denial is self-contradictory. For example "two is an even number" is a logical necessity.

Logical positivism   A school of philosophy of the first half of the twentieth century, aiming to combine empiricism and advances in logic to show all outstanding philosophical problems could be shown to be linguistic and solved by analysis of explication (see definition), or rational reconstruction of language. Logical positivists followed empiricists in holding that the only meaningful terms and statements refer to what experience can verify, whence their "verificationist criterion of meaningfulness."

Long-run relative frequency   An interpretation of probability according to which a statement of the probability of an outcome (say, tails on a coin flip) is equal to the total number of occurrences of the outcome (tails), divided by the total number of trials (all the coin flips), over the "long run," i.e. a run extended indefinitely into the future.

Metaphysics   The division of philosophy which examines the basic kinds of things there are in the universe. For example, the question "are there unobservable things?" is a metaphysical question. Compare **epistemology**.

Model   An intentionally simplified description of the regularities governing a natural process or a definition of such a system, usually mathematical and sometimes derived from a more general, less idealized or simplified theory, but sometimes developed independently of any theory. *See also* **semantic approach to theories**.

Natural kind   A metaphysical (see above) concept. By contrast with an artificial kind, a natural kind is a type of state, event, process or thing with existence independent of our classificatory interests. Thus, natural kinds are those which figure in natural laws (see below). "State capital" is an artificial kind, "acid" is a natural kind.

Natural law   A regularity that actually governs processes in nature and which science sets out to discover. Laws are usually thought to be of the conditional form, "If a then b" or "All As or Bs." Natural laws are often held to be true exceptionless regularities that underlie causal relations. *See* **scientific law**.

Naturalism   The philosophical thesis that the findings and methods of the natural sciences are the best guides to inquiry in philosophy, and particularly the philosophy of science. Naturalism rejects the claim that philosophy provides *a priori* foundations for science, and instead attempts to solve philosophical problems by exploiting theories in natural science. Naturalists are especially eager to derive insights for philosophy from Darwinian evolutionary theory.

Necessary condition   A condition whose absence presents an event from occurring or a statement from being true. For example, the presence of oxygen is a necessary condition for a match's lighting.

Necessary truth   A statement whose truth is not dependent on any contingent fact about the way the world just happens to be, but which reflects the only way things could be arranged. Contrast with contingent truth. For example, that 2 is an even number is a necessary truth.

Necessity *See* **logical necessity, physical necessity**.

Newtonian mechanics   The theory composed of the three laws of motion and the inverse square law of gravitation. This theory describes the behavior of all matter and is deterministic. Its success was the signal achievement of the seventeenth-century scientific revolution.

Normal science   The articulation of a paradigm, in which the scientist's task is to apply the paradigm to the solution of puzzles. Failure to solve puzzles is the fault of the scientists, not the paradigm. Persistent failure makes a puzzle an anomaly and threatens a revolution which may end the paradigm's hegemony.

Normative   Having to do with norms about the way things ought to be, as opposed to "positive" or "descriptive," having to do with the way things actually are, thus the realm of values, morality, ethics, policy.

Ontology   Metaphysics, the study of the basic kinds of things that exist. In the philosophy of science, more narrowly, the ontology of a theory are the kinds of things the theory is committed to the existence of. Thus, Newtonian mechanics is committed to the

existence of mass as an intrinsic property of things. Einsteinian mechanics is committed to mass as a relational property of things and their reference frames.

Paradigm   A term employed by Kuhn to characterize a scientific tradition, including its theory, textbook problems and solutions, its apparatus, methodology, and its philosophy of science. Paradigms govern normal science (see above). The term has come into general use to describe a world-view.

Partial interpretation   The thesis that observations give part of the meaning of theoretical terms.

Physical necessity   A statement is physically necessary if it is a law of nature or its truth follows from the laws of nature. Thus, it is physically necessary that no quantity of pure plutonium can have a mass of 100,000 kilograms, for the laws of physics tell us that long before it reached this mass, it would explode.

Positivism *See* **logical positivism**.

Pragmatics   The study of the contexts of communication which effect the meaning and success of an utterance. It is often held that the deductive–nomological model of explanation ignores the pragmatic dimensions along which we measure the success of an explanation requested and provided, in favor of purely non-pragmatic matters of logic and meaning.

Prior probability   In the Bayesian interpretation of probability, the prior probability is the betting odds assigned to a hypothesis before some new evidence is acquired that may change its probability via Bayes' theorem. According to Bayesianism, a scientist can begin with any assignment of a prior probability. Provided certain conditions obtain, so long as the scientist employs Bayes' theorem, the probabilities assigned to the hypothesis will eventually converge on the correct values.

Probabilistic propensity   The disposition some item has to exhibit some behavior with a certain frequency. For example, uranium atoms have the probabilistic propensity to emit alpha particles. Such propensities are mysterious because there is no underlying property of the systems which exhibit them that further explains the frequency of the behavior in question. Compare the disposition to be magnetic, which is explained by the orientation of electrons, or the disposition to be fragile, which is explained by chemical structure. Nothing explains a uranium atom's disposition to emit alpha particles with a certain frequency.

Probability   Either the subjective degree of belief that some proposition is true (Bayesian betting odds, see above) or the long-run relative frequency of something's happening under certain circumstances (weather-report probabilities), or the sheer likelihood that a given event will happen (probabilistic propensities in physics, see above). There are philosophical problems associated with each of these three definitions of probability.

Projectable   The property of a term or predicate that it names a natural kind (see above) and that the property can figure in natural laws. Coined by Goodman in his treatment of the problem of "grue" and "bleen."

Rational reconstruction *See* **explication**.

Rationalism   The epistemological thesis that in addition to or instead of sensory awareness, reason is a source of knowledge, either directly through logic or intuition.

Realism *See* **scientific realism; antirealism**.   The term is also employed to describe the position of Plato and his followers that numbers are real through abstract particular objects, and that properties, like being red or redness, exist independently of their instances—particularly red things.

Reduction   The relation between a less general and a more general theory in the same domain that enables the more general theory to explain the (approximate) truth of the less general theory, usually by logical derivation of the laws of the less general theory from the laws of the more general one. Thus, Newtonian mechanics is said to reduce Kepler's laws of planetary motion. Reduction will not obtain if theories are incommensurable (see above).

Relativism   The theory that nothing can be asserted to be true independent of some point of view, and that disagreements between points of view are irreconcilable, so that there are no unqualified truths.

Scientific law   Our best estimate as to a natural law. For example, Newton's inverse square law of gravitational attraction was for a long time held to describe an exceptionless regularity true everywhere and always, and therefore to constitute a natural law.

Scientific realism   The thesis that the claims of theoretical science must be treated as literally true or false, and that if we accept a theory as true, we are committed to the existence of its ontology (see above), the things it says there are, even if we cannot detect them. Compare **antirealism, instrumentalism**.

Semantic approach to theories   The claim that theories are not axiomatic systems (the syntactic approach, see below) but are sets of models, that is definitions of relatively simple systems with greater or lesser applicability to the world. The semantic approach is neutral on whether the models that constitute a theory reflect some underlying mechanism that explains their applicability.

Special sciences   A label invented by Fodor to describe those disciplines whose domains are fixed by local boundary conditions, including all the social and behavioral sciences, and possibly also biology.

Strong program   (in the sociology of science) The attempt to trace the nature of scientific change without relying on the fact that some theories are true or more approximately true than others. The program is motivated by the idea that since, as Kuhn has shown, there are losses as well as gains in scientific revolutions, and epistemic considerations cannot explain which theories triumph, the explanation of why they do so should appeal to factors no different from the factors which explain why some theories fail.

Sufficient condition   A condition whose presence guarantees the occurrence of an event or truth of a statement. For instance, being a son is a sufficient condition for being one's child.

Syntactic approach to scientific theories   The claim that theories are axiomatic systems in which empirical generalizations are explained by derivation from theoretical laws.

Synthetic truth   A statement true at least in part in virtue of contingent facts about the world. Thus, that "there are satellites circling Jupiter" is a synthetic truth. According to empiricism (see above), synthetic truths cannot be known *a priori*.

Teleological explanation   To explain some fact, event, process, state or thing by identifying the purpose, goal or end which it serves to attain. Since attaining a goal usually comes later, and sometimes does not obtain at all, such explanations do not appear to be causal, and are therefore suspect.

Testability   A statement is testable if definite consequences for observation can be inferred from it and compared to observations. Logical positivists demanded that all meaningful statements be testable. Post-positivist philosophers have accepted that no single statement is testable by itself.

Theory *See* **semantic approach to theories; syntactic approach to theories**.

Underdetermination   Theory is alleged to be underdetermined by data in that for any body of observational data, even all the observational data, more than one theory can be constructed to systematize, predict, and explain that data, so that no one theory's truth is determined by the data.

Verification   To establish the truth of a claim usually by observation. Positivists embraced a verificationist theory of meaning, according to which a statement was meaningful if and only if it was verifiable.

Bibliography

Achinstein, Peter, *Concepts of Science*, Baltimore, MD, Johns Hopkins University Press, 1967.

—— "Concepts of Evidence," *Mind* 87 (1978): 22–45.

—— *The Nature of Explanation*, Oxford, Oxford University Press, 1983.

—— "The Illocutionary Theory of Explanation," in Pitt, Joseph (ed.), *Theories of Explanation*, Oxford, Oxford University Press, 1988, 199–222.

Achinstein, Peter (ed.), *The Concept of Evidence*, New York, Oxford University Press, 1983.

Allen, C., Bekoff, M., and Lauder, G. (eds.), *Nature's Purposes: Analyses of Function and Design in Biology*, Cambridge, MA, MIT Press, 1998.

Anderson, E. "Feminist Epistemology: An Interpretation and Defense," *Hypatia* 10 (1995): 50–84.

Anthony, L., and Witt., C. (eds.), *A Mind of One's Own: Feminist Essays on Reason and Objectivity*, 2nd edition, Boulder, CO, Westview, 2001.

Ariew, A., Cummins, R., and Perlman, M. (eds.), *Functions: New Essays in the Philosophy of Psychology and Biology*, New York, Oxford University Press, 2002.

Ayer, A. J., "What Is a Law of Nature?" in *The Concept of a Person*, London, Macmillan, 1961. Reprinted in Curd, Martin, and Cover, Jan A. (eds.), *Philosophy of Science: The Central Issues*, New York, Norton, 1997, 808–825.

Balashov, Y., and Rosenberg, A. (eds.), *Philosophy of Science: Contemporary Readings*, London and New York, Routledge, 2002.

Barnes, Barry, *Scientific Knowledge and Social Theory*, London, Routledge, 1974.

Barnes, Barry, Bloor, David, and Henry, John, *Scientific Knowledge: A Sociological Analysis*, Chicago, University of Chicago Press, 1996.

Beauchamp, Tom L., and Rosenberg, Alex, *Hume and the Problem of Causation*, Oxford, Oxford University Press, 1981.

Bechtel, William, "Mechanism and Biological Explanation," *Philosophy of Science* 78(4) (2011): 533–557.

Becker, Adam, *What is Real? The Unfinished Quest for the Meaning of Quantum Physics*, New York, Basic Books, 2018.

Berkeley, George, *Principles of Human Knowledge*, first published 1710.

Bird, A., *Thomas Kuhn*, London, Acumen, 2000.

—— "The Dispositionalist Conception of Laws," *Foundations of Science* 10(4) (2005): 353–370.

Bloor, David, *Knowledge and Social Imagery*, London, Routledge, 1974.

Boyd, B., Gaspar, P., and Trout, J. D. (eds.), *The Philosophy of Science*, Cambridge, MA, MIT Press, 1991.

Braithwaite, Richard B., *Scientific Explanation*, Cambridge, Cambridge University Press, 1953.

Brandon, Robert, and McShea, Daniel, *Biology's First Law*, Chicago, University of Chicago Press, 2010.

Burtt, Edwin A., *The Metaphysical Foundations of Modern Science*, London, Routledge, 1926.

Butterfield, Herbert, *The Origins of Modern Science*, New York, Free Press, 1965.

Callender, C., and Cohen, J., "Better Best System Account of Lawhood in the Special Sciences," *Synthese* 28 (2008): 97–115.

Carnap, Rudolph, *The Continuum of Inductive Methods*, Chicago, University of Chicago Press, 1952.

Carroll, John (ed.), *Readings on Laws of Nature*, Pittsburgh, PA: Pittsburgh University Press, 2004.

Cartwright, Nancy, *How the Laws of Physics Lie*, Oxford, Oxford University Press, 1983.

Cartwright, Nancy, and Ward, Keith (eds.), *Rethinking Order: After the Laws of Nature*, London, Bloomsbury, 2016.

Chang, Hasok, *Is Water H_2O? Evidence, Realism, and Pluralism* (Boston Studies in the Philosophy and History of Science 293), New York, Springer, 2012.

Churchland, Paul, and Hooker, Clifford (eds.), *Images of Science: Essays on Realism and Empiricism*, Chicago, University of Chicago Press, 1985.

Cohen, I. Bernard, *The Birth of a New Physics*, New York, Norton, 1985.

Conant, James B. (gen. ed.), *Harvard Case Histories in the Experimental Sciences*, Cambridge, MA, Harvard University Press, 1957.

Craver, C., "Mechanisms and Natural Kinds," *Philosophical Psychology* 22(5) (2009): 575–594.

Craver, C., and Darden, D., *In Search of Mechanisms: Discoveries across the Life Sciences*, Chicago, University of Chicago Press, 2013.

Craver, C., and Kaplan, D. M., "The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective," *Philosophy of Science* 78(4) (2011): 601–627.

Curd, Martin, and Cover, Jan A. (eds.), *Philosophy of Science: The Central Issues*, New York, Norton, 1997.

Darwin, Charles, *On the Origin of Species*, New York, Avenel, 1979.

Davidson, D., *Inquiries into Meaning and Truth*, New York, Oxford University Press, 2001.

—— *Inquiries into Truth and Interpretation*, New York, Oxford University Press, 2001.

Dawkins, Richard, *The Blind Watchmaker*, New York, Norton, 1986.

Dennett, D., *Darwin's Dangerous Idea*, New York, Simon & Schuster, 1995.

Douglas, Helen, *Science, Policy, and the Value-Free Ideal*, Pittsburgh, PA, University of Pittsburgh Press, 2009.

Dretske, F., *Explaining Behavior: Reasons in a World of Causes*, Cambridge, MA, MIT Press, 1991.

Duhem, P., *The Aim and Structure of Physical Theory*, Princeton, NJ, Princeton University Press, 1991.

Earman, J., and Robert, J., "There Is No Problem of Provisos," *Erkenntnis* 57(3) (2002): 281–301.

Ellis, B., *Scientific Essentialism*, Cambridge, Cambridge University Press, 2007.

Epstein, Brian, *The Ant Trap: Rebuilding the Foundations of the Social Sciences*, Oxford, Oxford University Press, 2015.

Feyerabend, Paul, *Against Method*, London, Verso, 1975.

Feynman, Richard, *The Character of Physical Law*, Cambridge, MA, MIT Press, 1965.

—— *QED: The Strange Story of Light and Matter*, Princeton, NJ, Princeton University Press, 1984.

Fine, Arthur, "The Natural Ontological Attitude," in *The Shaky Game: Einstein, Realism, and Quantum Theory*, Chicago, University of Chicago Press, 1986, ch. 7.

Fodor, Jerry, "Special Sciences: Or the Disunity of Science as a Working Hypothesis," *Synthese* 28 (1974): 97–115.

—— *The Language of Thought*, New York: Thomas Crowell, 1975.

—— "Special Sciences: Still Autonomous After All These Years," in Tomberlin, J. E. (ed.), *Philosophical Perspectives*, New York, Blackwell, 1997, 149–164.

Frigg, R., "Models and Fiction," *Synthese* 172 (2010): 251–268.

Garson, Justin, "The Functional Sense of Mechanism," *Philosophy of Science* 80 (2013): 317–333.

Giere, R., *Explaining Science*, Chicago, University of Chicago Press, 1988.

—— *Science without Laws*, Chicago, University of Chicago Press, 1999.

Glymour, Clark, *Theory and Evidence*, Princeton, NJ, Princeton University Press, 1980.

Glymour, C., Spirtes, P., and Scheines, R., *Causation, Prediction, and Search*, 2nd edition, Cambridge, MA, MIT Press, 2001.

Godfrey-Smith, Peter, *Theory and Reality: An Introduction to the Philosophy of Science*, Chicago, University of Chicago Press, 2003.

—— "The Strategy of Model-Based Science," *Biology and Philosophy* 21 (2006): 725–740.

—— "Models and Fictions in Science," *Philosophical Studies* 143 (2009):101–116.

Goodman, Nelson, *Fact, Fiction and Forecast*, 3rd edition, Indianapolis, Bobbs-Merrill, 1973, first published 1948.

Greene, Brian, *The Elegant Universe*, New York, Vintage Books, 2000.

Gross, P., and Levitt, N., *Higher Superstition: The Academic Left and Its Quarrels with Science*, Baltimore, MD, Johns Hopkins University Press, 1994.

Gutting, Gary, *Paradigms and Revolutions*, Notre Dame, IN, University of Notre Dame Press, 1980.

Hacking, Ian, *An Introduction to Probability and Inductive Logic*, Cambridge, Cambridge University Press, 2001.

Harding, S., *The Science Question in Feminism*, Ithaca, NY, Cornell University Press, 1986.

Harding, S. (ed.), *The Feminist Standpoint Reader*, London, Routledge, 2003.

Harding, S., and O'Barr, J. F. (eds.), *Sex and Scientific Inquiry*, Chicago, University of Chicago Press, 1987.

Heilbron, J. L., *The History of Physics: A Very Short Introduction*, Oxford, Oxford University Press, 2018.

Hempel, Carl G., *Aspects of Scientific Explanation*, New York, Free Press, 1965.

—— "Empiricist Criteria of Significance: Problems and Changes," in *Aspects of Scientific Explanation*, New York, Free Press, 1965, 101–119.

—— "The Theoretician's Dilemma," in *Aspects of Scientific Explanation*, New York, Free Press, 1965, 173–228.

—— *Philosophy of Natural Science*, Englewood Cliffs, NJ, Prentice-Hall, 1966.

—— "Provisos," in Grunbaum, A., and Salmon, W. (eds.), *The Limitations of Deductivism*, Berkeley, University of California Press, 1988, 19–36.

Hoefer, C., and Rosenberg, A., "Empirical Equivalence, Underdetermination and Systems of the World," *Philosophy of Science* 61 (1994): 592–607.

Hofstadter, Douglas, *Gödel, Escher, Bach*, New York, Basic Books, 1999.

Horwich, Paul, *Probability and Evidence*, Cambridge, Cambridge University Press, 1982.

—— *World Changes: Thomas Kuhn and the Nature of Science*, Cambridge, MA, MIT Press, 1993.

Hull, David, *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*, Chicago, University of Chicago Press, 1988.

Hume, D., *A Treatise of Human Nature*, Oxford, Oxford University Press, 1888.

—— *Enquiry Concerning Human Understanding*, Indianapolis, Hackett Publishing Co., 1974.

Hypatia, Special Issue: "Analytic Feminism," 10(3) (1995): 1–182.

Janiak, A., *Newton as Philosopher*, Cambridge, Cambridge University Press, 2008.

Janiak, A. (ed.), *Newton: Philosophical Writings*, Cambridge, Cambridge University Press, 2014.

Jeffrey, Richard, *The Logic of Decision*, Chicago, University of Chicago Press, 1983.

Johnson, Gregory, *Argument and Inference: An Introduction to Inductive Logic*, Cambridge, MA, MIT Press, 2017.

Kant, Immanuel, *The Critique of Pure Reason*, London, Macmillan, 1961.

Kellert, S., Longino, H., and Waters, C. K., *Scientific Pluralism* (Minnesota Studies in the Philosophy of Science 19), Minneapolis, University of Minnesota Press, 2006.

Khalifa, Kareem, *Understanding, Explanation, and Scientific Knowledge*, Cambridge, Cambridge University Press, 2017.

Kim, J., *Physicalism or Something Near Enough*, Princeton, NJ, Princeton University Press, 2005.

Kitcher, Philip, *The Advancement of Science*, Oxford, Oxford University Press, 1995.

—— *Science in a Democratic Society*, Amherst, NY, Prometheus Books, 2011.

Kneale, William, *Probability and Induction*, Oxford, Oxford University Press, 1950.

Koertge, N. (ed.), *A House Built on Sand: Exposing Postmodernist Myths about Science*, New York, Oxford University Press, 1998.

Kuhn, Thomas, *The Copernican Revolution*, Cambridge, MA, Harvard University Press, 1957.

—— *The Essential Tension*, Chicago, University of Chicago Press, 1977.

—— *Black-Body Theory and the Quantum Discontinuity*, Chicago, University of Chicago Press, 1987.

—— *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press, 3rd edition, 1996.

—— *The Road since Structure*, edited by J. Conant and J. Haugeland, Chicago, University of Chicago Press, 2002.

Ladyman, J., and Ross, D., *Everything Must Go: Metaphysics Naturalized*, New York, Oxford University Press, 2009.

Lakatos, I., and Musgrave, A., *Criticism and the Growth of Knowledge*, Cambridge, Cambridge University Press, 1971.

Lange, M., *Natural Laws in Scientific Practice*, New York, Oxford University Press, 2000.

—— *Laws and Lawmakers*, Oxford, Oxford University Press, 2010.

—— *Because without Cause: Noncausal Explanations in Science and Mathematics*, Oxford, Oxford University Press, 2016.

Lange, M. (ed.), *Philosophy of Science: An Anthology*, Malden, MA, Blackwell, 2007.

Latour, Bruno, and Woolgar, Steven, *Laboratory Life: The Construction of Scientific Life*, London, Routledge, 1979.

Laudan, Larry, *Progress and Its Problems*, Berkeley, University of California Press, 1977.

Leibniz, G. W., *New Essays on Human Understanding*, Cambridge, Cambridge University Press, 1981.

Leplin, Jarrett, *A Novel Argument for Scientific Realism*, Oxford, Oxford University Press, 1998.

Leplin, Jarrett (ed.), *Scientific Realism*, Berkeley, University of California Press, 1984.

Leplin, J., and Laudan, L., "Empirical Equivalence and Underdetermination," *Journal of Philosophy* 88 (1991): 449–472.

Levins, R., and Lewontin, R., *The Dialectical Biologist*, Cambridge, MA, Harvard University Press, 1985.

Lewis, David, *Counterfactuals*, Oxford, Blackwell, 1974.

—— "Causation," in *Philosophical Papers*, vol. 2, Oxford, Oxford University Press, 1986, 159–214.

Lloyd, Elizabeth, *The Structure of Evolutionary Theory*, Princeton, NJ, Princeton University Press, 1987.

Locke, John, *Essay on Human Understanding*, first published 1690.

Loewer, B., "Why Is There Anything Except Physics?" *Synthese* 170 (2009): 217–233.

Longino, Helen, *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*, Princeton, NJ, Princeton University Press, 1990.

—— *The Fate of Knowledge*, Princeton, NJ, Princeton University Press, 2002.

Lycan, William, *The Philosophy of Language*, 2nd edition, London, Routledge, 2008.

MacFarlane, J., *Assessment Sensitivity: Relative Truth and its Applications*, Oxford, Oxford University Press, 2014.

Mach, Ernst, *The Analysis of Sensation*, first published 1886.

Machamer, P., Darden, L., and Craver, C. F., "Thinking about Mechanisms," *Philosophy of Science* 67 (2000): 1–25,

Mackie, John L., *Truth, Probability, and Paradox: Studies in Philosophical Logic*, Oxford, Oxford University Press, 1973.

—— *The Cement of the Universe*, Oxford, Oxford University Press, 1974.

Maudlin, Tim, *The Metaphysics within Physics*, Oxford, Clarendon Press, 2007.

Mayo, Deborah, *Error and the Growth of Experimental Knowledge*, Chicago, University of Chicago Press, 1996.

McIntyre, Lee, *The Scientific Attitude: Defending Science from Denial, Fraud, and Pseudoscience*, Cambridge, MA, MIT Press, 2019.

McIntyre, L., and Rosenberg, A. (eds.), *The Routledge Companion to Philosophy of Social Science*, New York, Routledge, 2016.

McShea, D., and Rosenberg, A., *Philosophy of Biology: A Contemporary Approach*, London and New York, Routledge, 2007.

Merton, Robert K., *The Sociology of Science*, Chicago, University of Chicago Press, 1973.

Mill, John S., *A System of Logic*, first published 1843.

Miller, Richard, *Fact and Method: Explanation, Confirmation and Reality in the Natural Sciences*, Princeton, NJ, Princeton University Press, 1987.

Morgan, Mary S., and Morrison, Margaret (eds.), *Models as Mediators: Perspectives on Natural and Social Science*, Cambridge, Cambridge University Press, 1999.

Nagel, Ernest, *Teleology Revisited*, New York, Columbia University Press, 1977.

—— *The Structure of Science*, 2nd edition, Indianapolis, Hackett, 1979.

Nagel, Ernest, and Newman, James R., *Gödel's Proof*, New York, New York University Press, 1954.

—— *Gödel's Proof*, new edition, New York, New York University Press, 2008.

Nelson, L., *Who Knows: From Quine to a Feminist Empiricism*, Philadelphia, PA, Temple University Press, 1992.

Newton-Smith, William, *The Rationality of Science*, London, Routledge, 1981.

Nickles, Thomas, *Thomas Kuhn*, Cambridge, Cambridge University Press, 2002.

Okasha, Samir, *Philosophy of Science: A Very Short Introduction*, Oxford, Oxford University Press, 2002.

Park, K., and Daston, L. (eds.), *The Cambridge History of Science*, vol. 3: *Early Modern Science*, Cambridge, Cambridge University Press, 2006.

Paul, L., and Hall, N., *Causation: A User's Guide*, Oxford, Oxford University Press, 2013.

Pearl, Judea, *Causality: Models, Reasoning, and Inference*, 2nd edition, Cambridge, Cambridge University Press, 2009.

Pearl, Judea, and Mackenzie, Dana, *The Book of Why: The New Science of Cause and Effect*, New York, Basic Books, 2018.

Piccinini, G., and Craver, C., "Integrating Psychology and Neuroscience: Functional Analysis as Mechanism Sketches," *Synthese* 183 (2011): 283–311.

Pickering, Andrew, *Constructing Quarks*, Chicago, University of Chicago Press, 1984.

Pigliucci, M., and Boudry, M. (eds.), *Philosophy of Pseudoscience: Reconsidering the Demarcation Problem*, Chicago, University of Chicago Press, 2013.

Pinnick, C., Koertge, N., and Almeder, R. (eds.), *Scrutinizing Feminist Epistemology: An Examination of Gender in Science*, New Brunswick, NJ, Rutgers University Press, 2003.

Pitt, Joseph (ed.), *Theories of Explanation*, Oxford, Oxford University Press, 1988.

Popper, Karl R., *The Logic of Scientific Discovery*, London, Hutchinson, 1959, first published in German in 1935.

—— *Objective Knowledge*, New York, Harper and Row, 1984.

Porter, Roy (ed.), *The Cambridge History of Science*, vol. 4: *The Eighteenth Century*, Cambridge, Cambridge University Press, 2003.

Quine, Willard V. O., *From a Logical Point of View*, Cambridge, MA, Harvard University Press, 1951.

—— *Word and Object*, Cambridge, MA, MIT Press, 1961.

Railton, Peter, "A Deductive-Nomological Model of Probabilistic Explanation," in Pitt, Joseph (ed.), *Theories of Explanation*, Oxford, Oxford University Press, 1988, 199–222.

Ramsey, Frank, "The Foundations of Mathematics," in *The Foundations of Mathematical Logic and Other Logical Essays*, London, Routledge & Kegan Paul, 1925, 1–61.

Reichenbach, Hans, *Experience and Prediction*, Chicago, University of Chicago Press, 1938.

—— *The Rise of Scientific Philosophy*, Berkeley, University of California Press, 1951.

Rosenberg, Alex, *The Structure of Biological Science*, Cambridge, Cambridge University Press, 1985.

—— *Philosophy of Social Science*, Boulder, CO, Westview, 1992.

Ruse, Michael, *But is it Science? The Philosophical Question in the Creation/Evolution Controversy*, Amherst, NY, Prometheus Books, 1988.

Salmon, Wesley C., *Foundations of Scientific Inference*, Pittsburgh, PA, University of Pittsburgh Press, 1966.

—— *Scientific Explanation and the Causal Structure of the World*, Princeton, NJ, Princeton University Press, 1984.

—— "Statistical Explanation and Causality," in Pitt, Joseph (ed.), *Theories of Explanation*, Oxford, Oxford University Press, 1988, 75–119.

—— "Four Decades of Scientific Explanation," in Salmon, Wesley, and Kitcher, Philip (eds.), *Scientific Explanation* (Minnesota Studies in the Philosophy of Science 13), Minneapolis, University of Minnesota Press, 1989.

Salmon, Wesley, and Kitcher, Philip (eds.), *Scientific Explanation* (Minnesota Studies in the Philosophy of Science 13), Minneapolis, University of Minnesota Press, 1989.

Savage, Leonard, *Foundations of Statistics*, New York, Dover, 1972.

Scerri, Eric, *A Tale of Seven Elements and a New Philosophy of Science*, Oxford, Oxford University Press, 2016.

Schilpp, P. A., *Albert Einstein: Philosopher-Scientist*, Evanston, IL, Open Court, 1949.

Shapere, Dudley, "Review of *The Structure of Scientific Revolutions*," *Philosophical Review* 73 (1964): 383–394.

Shapin, Steven, *The Scientific Revolution*, Chicago, University of Chicago Press, 1998.

Shapin S., and Schaffer, S., *Leviathan and the Air Pump: Hobbes, Boyle, and the Experimental Life*, Princeton, NJ, Princeton University Press, 2011.

Sheffler, Israel, *Science and Subjectivity*, Indianapolis, Bobbs-Merrill, 1976.

Skyrms, B., *From Zeno to Arbitrage: Essays on Quantity, Coherence, and Induction*, New York, Oxford University Press, 2013.

Smart, J. J. C., *Between Science and Philosophy*, London, Routledge, 1968.

Soames, S., *Philosophical Analysis in the Twentieth Century*, 2 vols., Princeton, NJ, Princeton University Press, 2005.

Sober, E., *The Nature of Selection*, Cambridge, MA, MIT Press, 1984.

—— *Philosophy of Biology*, 2nd edition, Boulder, CO, Westview Press, 1999.

Sokal, A., and Bricmont, J., *Intellectual Impostures*, London, Profile, 1998.

Spector, Marshall, *Concepts of Reduction in Physical Science*, Philadelphia, PA, Temple University Press, 1968.

Stanford, P. K., *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*, New York, Oxford University Press, 2010.

Sterelny, K., and Griffiths, P., *Sex and Death*, Chicago, University of Chicago Press, 1997.

Stove, David C., *Hume, Probability, and Induction*, Oxford, Oxford University Press, 1967.

Strevens, Michael, *Depth: An Account of Scientific Explanation*, Cambridge, MA, Harvard University Press, 2008.

Suppe, Fredrick, *The Structure of Scientific Theories*, Urbana, University of Illinois Press, 1977.

Swinburne, R., *Bayes' Theorem*, Oxford, Oxford University Press, 2005.

Swoyer, C., "The Nature of Natural Laws," *Australasian Journal of Philosophy* 60(3) (1982): 203–223.

Thompson, Paul, *The Structure of Biological Theories*, Albany, NY, SUNY Press, 1989.

Tooley, Richard M., *Causation: A Realist Approach*, Oxford, Oxford University Press, 1987.

van Fraassen, Bas, *The Scientific Image*, Oxford, Oxford University Press, 1980.

—— "The Pragmatic Theory of Explanation," in Pitt, Joseph (ed.), *Theories of Explanation*, Oxford, Oxford University Press, 1988, 136–155.

Weinberg, S., *Dreams of a Final Theory*, New York, Random House, 1993.

Weisberg, M., and Matthewson, J., "The Structure of Tradeoffs in Model Building," *Synthese* 170 (2009): 169–190.

Weiskopf, D., "Models and Mechanisms in Psychological Explanation," *Synthese* 183 (2011): 313–338.

Westfall, Richard, *The Construction of Modern Science*, Cambridge, Cambridge University Press, 1977.

—— *Never at Rest*, Cambridge, Cambridge University Press, 1983.

Whitehead, A. N., *Science and the Modern World*, New York, Free Press, 1997.

Wilson, E. O., *Consilience: The Unity of Knowledge*, New York, Knopf, 1998.

Woodward, J., "What is a Mechanism? A Counterfactual Account," *Philosophy of Science* 69(S3) (2002): S366–S377.

—— *Making Things Happen*, Oxford, Oxford University Press, 2003.

Wray, K. Brad, *Kuhn's Evolutionary Social Epistemology*, Cambridge, Cambridge University Press, 2014.

Wright, Larry, *Teleological Explanation*, Berkeley, University of California Press, 1976.

Index