an introduction to the philosophy of science

THEORY AND REALITY

PETER GODFREY-SMITH

The University of Chicago Press / Chicago and London

Peter Godfrey-Smith is associate professor of philosophy and of history and philosophy of science at Stanford University. He is the author of *Complexity and the Function of Mind in Nature*.

The University of Chicago Press, Chicago 60637
The University of Chicago Press, Ltd., London

Printed in the United States of America

12 11 10 09 08 07 06 05 04 03 1 2 3 4 5

ISBN: 0-226-30062-5 (cloth)
ISBN: 0-226-30063-3 (paper)

Library of Congress Cataloging-in-Publication Data

Godfrey-Smith, Peter.
Theory and reality : an introduction to the philosophy of science / Peter Godfrey-Smith.
p. cm. — (Science and its conceptual foundations)
Includes bibliographical references and index.
ISBN 0-226-30062-5 (alk. paper)— ISBN 0-226-30063-3 (pbk. : alk. paper)
1. Science—Philosophy. I. Title. II. Series.
Q175 .G596 2003
501—dc21
2002155305

⊚ The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992.

In any case, this chapter and the previous one have introduced some of the main themes in naturalistic philosophy of science. Naturalists hope that by combining philosophical analysis with input from other disciplines, we will eventually get a complete picture of how science works and what sort of connection it gives us to the world. This last issue—the connection that science gives us to the real world we inhabit—has often been mishandled by sociology of science and Science Studies. That is the topic of the next chapter.

Further Reading

For assessments of Hull's theory of science, see the reviews in *Biology and Philosophy,* volume 3 (1988). Also see Sterelny 1994.

Kitcher's main work here is *The Advancement of Science* (1993). The model discussed in this chapter is presented in a simpler form in Kitcher 1990. Solomon (2001) defends a "social empiricism" in detail, with many examples from the history of science.

For a more general discussion of social structure and epistemology, see Goldman 1999. Downes (1993) argues that some naturalists do not take the social nature of science seriously enough. Sulloway 1996 is a very adventurous discussion of the role of personality and temperament in scientific revolutions.

12

Scientific Realism

12.1 Strange Debates

What does science try to describe? The world, of course. Which world is that? *Our* world, the world we all live in and interact with. Unless we have made some very surprising mistakes in our current science, the world we now live in is a world of electrons, chemical elements, and genes, among other things. Was the world of one thousand years ago a world of electrons, chemical elements, and genes? Yes, although nobody knew it back then.

But the concept of an electron is the *product* of debates and experiments that took place in a specific historical context. If someone said the word "electron" in 1000 A.D., it would have meant nothing—or at least certainly not what it means now. So how can we say that the world of 1000 A.D. was a world *of* electrons? We cannot; we must instead regard the existence of electrons as dependent on our conceptualization of the world.

Those two paragraphs summarize one part of an argument about science that has gone on constantly for the last fifty years, and which stretches much deeper into the history of philosophy. For some people, the claims made in the first paragraph are so obvious that only a tremendously confused person could deny them. The world is one thing, and our ideas about it are another! For other people, the arguments in the second paragraph show that there is something badly wrong with the simple-looking claims in the first paragraph. The idea that our theories describe a real world that exists wholly independently of thought and perception is a mistake, a naive philosophical view linked to other mistakes about the history of science and the place of science in society.

These problems have arisen several times in this book. In chapter 6 we looked at Kuhn's claim that when paradigms change, the world changes too. In chapter 8 we found Latour suggesting that nature is the "product" of the settlement of scientific controversy. I criticized those claims, but now it is time to give a more detailed account of how theory and reality are connected.

12.2 Approaching Scientific Realism

The position defended in this book is a version of *scientific realism*. A scientific realist thinks it *does* make sense to say that science aims at describing the real structure of the world we live in. Does the scientific realist think that science *succeeds* in this aim? That is a more complicated issue.

Formulating scientific realism in a precise way will take a while. And the best way to start is to ignore science for the moment and look first for a more general description of "realist" attitudes.

The term "realism" gets used in a huge variety of ways in philosophy; this is a term to be very cautious about. One tradition of dispute has to do with what our basic attitude should be toward the world that we seem to inhabit. The simple, common-sense view is that the world is out there around us, existing regardless of what we think about it. But this simple idea has been challenged over and over again. One line of argument holds that we could never *know* anything about a world of that kind. This debate has carried over into the philosophy of science.

How might we give a more precise formulation of the "common-sense" realist position? The usual starting point is the idea that reality is "independent" of thought and language (Devitt 1997). This idea is on the right track, but it has to be understood carefully. People's thoughts and words are, of course, real *parts* of the world, not extra things floating somehow above it. And thought and language have a crucial *causal* role in the world. One of the main reasons for thinking, talking, and theorizing is to work out how to affect and transform things around us. Every bridge or light bulb is an example of this phenomenon. So a realist statement about the independence of the world from thought must have some qualifications. Here is my formulation:

Common-sense Realism: We all inhabit a common reality, which has a structure that exists independently of what people think and say about it, except insofar as reality is comprised of, or is causally affected by, thoughts, theories, and other symbols.

The realist accepts that we may all have different *views* about the world and different perspectives on it. Despite that, we are all here living in and interacting with the same world. Let us now return to issues involving science.

12.3 A Statement of Scientific Realism

How should scientific realism be formulated? One possibility is to see the scientific realist as asserting that the world really is the way it is described by our best-established scientific theories. We might say: there really are electrons, chemical elements, genes, and so on. The world as described by science is the real world. Michael Devitt is an example of a scientific realist who expresses his position in this way (1997).

My approach will be different. I agree with Bas van Fraassen, and others, who argue that it is a mistake to express the scientific realist position in a way that depends on the accuracy of our current scientific theories. If we express scientific realism by asserting the real existence of the entities recognized by science now, then if our current theories turn out to be false, scientific realism will be false too.

Should we worry about the possibility that our best-established theories will turn out to be wrong? Devitt thinks that so long as we do not commit ourselves to realism about speculative ideas at the frontiers of science, we need not worry. Others think that this confidence shows disregard for the historical record; we should always recognize the genuine possibility that well-established parts of science will run into trouble in the future.

How do we decide this massive question about the right level of confidence to have in current science? My suggestion is that we *don't* decide it here. Instead we should separate this question from the question of scientific realism. A scientific realist position is compatible with a variety of different attitudes about the reliability of our current theories. We want a formulation of scientific realism that is expressed as a claim about the enterprise of science as a whole.

One complication comes from the following question: must the scientific realist also be a common-sense realist? Is it possible—in principle—that science could tell us that common-sense realism is false? The problem is made vivid by puzzles with quantum mechanics, one of the basic theories in modern physics. According to quantum mechanics, the state of a physical system is partially determined by the act of measurement. Some interpretations of quantum mechanics see this as causing problems for common-sense realist ideas about the relation between human thought and physical reality. These interpretations of quantum mechanics are very controversial. Like a lot of other philosophers, I have been quietly hoping that further work will eventually show them to be completely mistaken. But that is not the point that matters here. The point is this: should we allow for the *possibility* that science could conflict with common-sense realism? If we say that scientific realism does assume common-sense realism, we seem to be committed to holding on to an everyday, unreflective picture of the world, regardless of what science ends up saying. But if we sever scientific realism from common-sense realism, it becomes hard to formulate a general claim about how the aim of science is to describe the real world.

My response to the problem is to modify common-sense realism so that it allows for the possibility of unexpected, uncommonsensical relations between thought and reality at large. Common-sense realism as previously formulated allowed for the possibility of *causal* links between thought and the rest of reality. It is often hard to tell whether a connection posited by science is a causal connection or not. So let us widen the class of relations between thought and the world that realism accepts; science might add new cases. Because we are modifying common-sense realism to make it more responsive to science, this is a *naturalistic* modification.

Common-sense Realism Naturalized: We all inhabit a common reality, which has a structure that exists independently of what people think and say about it, except insofar as reality is comprised of thoughts, theories, and other symbols, and except insofar as reality is dependent on thoughts, theories, and other symbols in ways that might be uncovered by science.

Once we have made this modification, it is reasonable to include common-sense realism as part of scientific realism. Here is my preferred statement of scientific realism:

Scientific Realism:
1. Common-sense realism naturalized.
2. One actual and reasonable aim of science is to give us accurate descriptions (and other representations) of what reality is like. This project includes giving us accurate representations of aspects of reality that are unobservable.

In this sense, I am a scientific realist.

Several comments on this formulation are needed. First, clause 2 says that *one* aim of science is to represent the structure of the world. Nothing implies here that this is the only aim of science. There might be other aims as well. And some particular theories—even whole research programs—might be developed in a way intended to serve other purposes.

Second, I said "actual and reasonable aim." The first part of this is a claim about the goals behind at least a good proportion of actual scientific work. The second part claims that scientists are not deluded or irrational in making this their goal. They can reasonably hope to succeed at least some of the time.

But I do not say how often they succeed. No part of my statement of scientific realism endorses our current particular scientific theories. In some areas of science, it's hard to imagine that we could be badly wrong in our current views; it's hard to imagine that we could be wrong in believing that

tuberculosis is caused by a bacterium and that chemical bonding occurs via the interactions of outer-shell electrons in atoms. Still, my statement of scientific realism is intended to capture the possibility of both *optimistic* and *pessimistic* versions. An optimistic scientific realist thinks we can be confident that science is succeeding in uncovering the basic structure of the world and how it works. The pessimistic option is more cautious, even slightly skeptical. A pessimistic scientific realist might be someone who thinks that it is very hard for our feeble minds to get to the right theories, that evidence is often misleading, and that we tend to get too confident too quickly.

So there is a range of possible attitudes within scientific realism toward our chances of really understanding how the world works. Although there is a range, there is also a limit. My statement of scientific realism says that giving us accurate representations of the world is a *reasonable* aim of science. If someone thought it was just about *impossible* for us to get to the right theories, then it is hard to see how it could be a reasonable aim of science to *try* to do so. So there is a limit to the pessimism that is compatible with scientific realism as I understand it; *extreme* pessimism is not compatible. I think of Popper as someone who is getting close to this limit but who does not actually reach it.

Although Kuhn's most famous discussions of realism are his notorious claims about how the world changes when paradigms change, at other times he seems more like a pessimistic scientific realist. These are passages where Kuhn seems to think that the world is just so *complicated* that our theories will always run into trouble in the end—and this is a fact about the world that is independent of paradigms. We try to "force" nature into "boxes," but nature resists. All paradigms are doomed to fail eventually. This skeptical realist view is more coherent and more interesting than Kuhn's "changing worlds" position.

Much of the recent philosophical debate under the heading "scientific realism" has really been discussion of whether we should be optimistic or pessimistic about the aspirations of science to represent the world accurately (Psillos 1999). Some hold that fundamental ideas have changed so often within science—especially within physics—that we should always expect our current views to turn out to be wrong. Sometimes this argument is called the "pessimistic meta-induction." The prefix "meta" is misleading here, because the argument is not an induction about inductions; it's more like an induction about explanatory inferences. So let's call it "the pessimistic induction from the history of science." The pessimists give long lists of previously posited theoretical entities like phlogiston and caloric that we now think do not exist (Laudan 1981). Optimists reply with long lists of theoretical entities that once were questionable but which we now think definitely do exist—like atoms, germs, and genes.

These debates only have the ability to threaten scientific realism of the kind defended here it if they threaten to establish *extreme* pessimism. They do not support extreme pessimism. But the debates are interesting in their own right. What level of confidence should we have in our current theories, given the dramatic history of change in science? We should not think that this question is one to be settled *solely* by the historical track record. We might have reason to believe that our methods of hypothesizing and testing theories have improved over the years. But history will certainly give us interesting data on the question.

We might find good reason to have different levels of confidence, and also different *kinds* of confidence, in different domains of science. Ernan McMullin (1984) has rightly urged that we not think of the parts of physics that deal with the ultimate structure of reality as a model for all of science. Basic physics is where we deal with the most inaccessible entities, those furthest from the domain our minds are adapted to dealing with. In basic physics we often find ourselves with powerful mathematical formalisms that are hard to interpret. These facts give us grounds for caution. And where we are optimistic, we might have grounds for optimism about some features of our theories and not others. McMullin and also John Worrall (1989) have developed versions of the idea that the confidence we should have about basic physics is confidence that low-level *structural* features of the world have been captured reliably by our models and equations. That is a special *kind* of confidence.

All those factors that are relevant in the case of fundamental physics *do not apply* in the case of molecular biology. There we deal with entities that are far from the lowest levels, entities that we have a variety of kinds of access to. We do not find ourselves with powerful mathematical formalisms that are hard to interpret. The history of this field also supports a view holding that we are steadily accumulating knowledge of how biological molecules work and how they operate in the processes of life. So trying to work out the right attitude to have toward molecular biology is not the same as trying to work out the right attitude toward theoretical physics.

Realists sometimes claim that there is a general argument from the *success* of scientific theories to their truth. It is sometimes claimed that realism is the only philosophy of science that does not make the success of science into a miracle (Smart 1968; Putnam 1978). This line of argument has been unimpressive as a defense of realism. The real world will definitely have *some* role in affecting the success and failure of theories. Theories will do well or badly *partly* because of their relations to the world in which they

are used and investigated. But there are *many* kinds of ways in which the link between theory and reality can generate success, especially in the short or medium term. Accurate representation of the world is not the only way. Theories can contain errors that compensate for each other. And theories can be successful despite being very wrong about the *kinds of things* they posit, provided they have the right *structure* in crucial places. Here is a simple example used by Laudan: Sadi Carnot thought that heat was a fluid, but he worked out some of the basic ideas of thermodynamics accurately despite this. The flow of a fluid was similar enough to patterns in the transfer of kinetic energy between molecules for his mistake not to matter much. Realists need to give up the idea that success in science points directly or unambiguously toward the truth of theories.

I hope my reasons for setting things up in the way I have are becoming clear. Much of the literature has held that scientific realists must be optimistic about current theories and about the history of science. I resisted that formulation of the issue. There is no point in arguing too much about the term "scientific realism," but there are benefits from organizing the issues in the way I have. What I call scientific realism is a fairly definite yes-or-no choice. (*Fairly* definite; see section 12.7.) This is also a choice about fundamental philosophical issues. The question about the right level of optimism to have about well-established scientific theories is *not* a question that has a simple answer that can be easily summarized. There we need to distinguish between different scientific fields, different kinds of theories, different kinds of success, and different kinds of optimism. In many cases we surely have good reason to be optimistic, but simple slogans should not be trusted.

One more comment on my formulation of scientific realism is needed. I said that science aims to give us "accurate descriptions and other representations of what reality is like." This is meant to be very broad, because there are lots of different kinds of representation used by different sciences. Some philosophers think that the main goal for a realist is *truth;* a good theory is a true theory. So they might want to formulate realism by saying that science aims to give us true theories. But the concepts of truth and falsity are only easy to apply in cases where a representation is in the form of language. In addition to linguistic representations, science often uses mathematical models, and other kinds of models, to describe phenomena. A scientific claim might also be expressed using a diagram. So I use the term "accurate representation" in a broad way to include true linguistic descriptions, pictures and diagrams that resemble reality in the way they are supposed to, models that have the right structural similarity to aspects of the world, and so on. I will return to these issues in the final section of this chapter.

12.4 Challenges from Traditional Empiricism

Scientific realism is now a popular position, but it has faced constant criticisms and challenges. Many of the most influential philosophers have thought that there is at least *something* wrong with scientific realism of the kind described in the previous section. Let's do a quick survey of the philosophers discussed so far in this book. Logical positivism was mostly opposed to scientific realism. Kuhn was vague and not always consistent, but he mostly opposed it. Many sociologists of science have certainly opposed it, including Latour. Goodman, the inventor of the "new riddle of induction," was opposed to it. Van Fraassen, who influenced my statement of what scientific realism is, rejects the view. So does Laudan. Feyerabend is hard to assess. Popper is in favor of scientific realism. Many of the naturalists discussed in the two previous chapters are scientific realists (including Fodor, Hull, and Kitcher), but not all are.

The critics listed above do not agree on *what* is wrong with scientific realism. I will divide the various forms of opposition into three broad families. Critics of realism differ among themselves just as much as they differ from the realists.

First, scientific realism has often been challenged by traditional forms of empiricism. In this book I will defend both scientific realism *and* a kind of empiricism, but this is not always an easy alliance. Indeed, one side of the debate about realism is often referred to as a debate *between* realism and empiricism.

Traditional empiricists tend to worry about both common sense and scientific realism, and they often worry for reasons having to do with knowledge. If there was a real world existing beyond our thoughts and sensations, how could we ever know anything about it? Empiricists believe that our senses provide us with our only source of factual knowledge. Many empiricists have thought that sensory evidence is not good enough for us to regard ourselves as having access to a "real world" of the kind the realist is committed to. And it seems strange (though not absurd, I think) to be in a position where you simultaneously say that a real world exists and also say we can never have any knowledge about it whatsoever.

The logical positivists recast these issues in terms of their theory of language. In the heyday of logical positivism, traditional philosophical questions about the "reality of the external world" were regarded as meaningless and empty. So the logical positivist attitude to most discussions of the "relation between science and reality" is that no side of the debate is saying anything meaningful and the whole discussion is a waste of time.

Some versions of logical positivism were also committed to the "phenomenalist" idea that all meaningful sentences can be translated into sentences that refer only to sensations. If phenomenalism is true, then when we *seem* to make claims about real external objects, all we are talking about are patterns in our sensations. Some more holistic empiricist views about language, of the kind associated with logical empiricism, have the same consequence. Even if translations are not possible, the nature of language prevents us from hoping to describe the structure of a world beyond our senses. Language and thought just cannot "reach" that far. I believe that a lot of twentieth-century empiricism held onto a version of this view (though some commentators on this book have objected to this claim).

In recent years the tension between realism and empiricism has often been debated under the topic of the "underdetermination of theory by evidence." Empiricists argue that there will always be a range of alternative theories compatible with all our actual evidence, and maybe a range of alternative theories compatible with all our possible evidence. So we never have good empirical grounds for choosing one of these theories over others and regarding it as representing how the world really is. This takes us back to the discussion in the previous section about the right level of optimism we should have about our scientific theories. I expressed scientific realism in a way compatible with a fair degree of pessimism, but the problem of underdetermination is important in its own right (see also sections 15.2 and 15.3).

12.5 Metaphysical Constructivism

I use the term "metaphysical constructivism" for a family of views including those of Kuhn and Latour. These views hold that, in some sense, we have to regard the world as *created* or *constructed* by scientific theorizing. Kuhn expressed this claim by saying that when paradigms change, the world changes too. Latour expresses the view by saying that nature (the real world) is the *product* of the decisions made by scientists in the settlement of controversies. Nelson Goodman is another example; he argues that when we invent new languages and theories, we create new "worlds" as well (1978). For a metaphysical constructivist, it is not even *possible* for a scientific theory to describe the world as it exists independent of thought, because reality itself is dependent on what people say and think.

These views are always hard to interpret, because they look so strange when interpreted literally. How could we possibly *make* the world just by making up a new theory? Maybe Kuhn, Latour, and Goodman are just using a metaphor of some kind? Perhaps. Kuhn sometimes expressed a different view on the question, a kind of skeptical realism, and he struggled to make his position clear. But when writers such as Goodman have been asked about this, they have generally insisted that their claims are *not* just metaphorical (Goodman 1996, 145). They think there is something quite wrong with scientific realism of the kind I described in section 12.3. They accept that it's hard to describe a good alternative, but they think we should use the concept of "construction," or something like it, to express the relationship between theories and reality.

Some of these ideas can be seen as modified versions of the view of Immanuel Kant ([1781] 1998). Kant distinguished the "noumenal" world from the "phenomenal" world. The noumenal world is the world as it is in itself. This is a world we are bound to believe in, but which we can never know anything about. The phenomenal world is the world as it appears to us. The phenomenal world is knowable, but it is partly our creation. It does not exist independently of the structure of our minds.

This kind of picture has often seemed appealing to philosophers who want to deny scientific realism but do so in a moderate way. Hoyningen-Huene (1993) has argued that we should interpret Kuhn's views as similar to Kant's. In Michael Devitt's analysis of the realism debates (1997), a wide range of philosophers are seen as either deliberately or inadvertently following the Kantian pattern. According to Devitt, constructivist antirealism works by combining the Kantian picture with a kind of relativism, with the idea that different people or communities create different "phenomenal worlds" via the imposition of their different concepts on experience. This relativist idea was not part of Kant's original view; for Kant all humans apply the same basic conceptual framework and have no choice in the matter.

The Kantian picture is sometimes seen as a way of holding onto the idea that there is a real world *constraining* what we believe but doing so in a way that does not permit our knowing or representing this world. This move is often tempting, but the resulting views are unhelpful. Understanding our access to reality is difficult, but adding an extra layer called "the phenomenal world" in between us and the real world achieves nothing.

The term "social constructivism" is often used for roughly the same kind of view that I am calling metaphysical constructivism. But "social constructivism" is also used for more moderate ideas as well. If someone argues that we make or construct our *theories,* or our classifications of objects, that claim is not opposed to scientific realism. We do indeed "construct" our ideas and classifications. Nature does not hand them to us on a platter. But a scientific realist insists that beyond ideas and theories there is also the rest of reality.

In fields like sociology of science, as we saw in chapter 8, there is an unfortunate tradition of not explicitly distinguishing between the construction of ideas and the construction of reality. What is it about these fields that has encouraged such strange-sounding formulations of ideas? There are various reasons, but I will venture some meta-sociology here—sociology of the sociology of science. A lot of work in these fields has been organized around the desire to oppose a particular *Bad View* that is seen as completely wrong. The Bad View holds that reality determines thought by stamping itself on the passive mind; reality acts on scientific belief with "unmediated compulsory force" (Shapin 1982, 163). That picture is to be avoided at all costs; it is often seen as not only false but even *politically* harmful, because it suggests a passive, inactive view of human thought. Many traditional philosophical theories are interpreted as implicitly committed to this Bad View. This is one source for descriptions of logical positivism as reactionary, helpful to oppressors, and so on.

What results from this is a tendency for people to go *as far as possible* away from the Bad View. This encourages people to assert simple *reversals* of the Bad View's relationship between mind and world. Thus we reach the idea that theories construct reality.

Some explicitly embrace the idea of an "inversion" of the traditional picture (Woolgar 1988, 65), while others leave things more ambiguous. But there is little pressure within the field to discourage people from going too far in these statements. (Bloor 1999 is an interesting exception.) Indeed, those who express more moderate denials of the Bad View leave themselves vulnerable to criticism from within the field. The result is a literature in which one error—the view that reality stamps itself on the passive mind—is exchanged for another error, the view that thought or theory constructs reality.

12.6 Van Fraassen's View

The last form of opposition to scientific realism that I will discuss is a more moderate and careful form; this is the position of Bas van Fraassen (1980). Van Fraassen's ideas lie within the empiricist tradition, but they are not based on a linguistic or psychological theory. Instead, van Fraassen confronts realism on the proper aims of science. So his antirealism is a direct denial of the kind of scientific realism defended in this chapter. This is no accident, since my formulation of scientific realism was influenced by his.

In discussions of realism, the term "instrumentalism" is used to refer to a variety of antirealist views. Sometimes it is used for traditional empiricist positions of the kind discussed earlier. But sometimes it is used in a different way, which I think is more appropriate. According to instrumentalism in this sense, we should think of scientific theories as devices for helping us deal with experience. Rather than saying that describing the real world is impossible, an instrumentalist will urge us *not to worry* about whether a theory is a true description of the world, or whether electrons "really, really exist." If a theory enables us to make good predictions, what more can we ask? If we have a theory that gives us the right answers with respect to what we can observe, we might occasionally find ourselves wondering if these right answers result from some deeper "match" between the theory and the world. But we can never expect to know the answer to this question, so what relevance does it have to science? Quite a few scientists have expressed instrumentalist views, especially in physics. The idea that we should ignore questions about the "real reality" of theoretical entities because these questions have no practical relevance is also linked to one strand of the *pragmatist* tradition in philosophy (Rorty 1982).

A detailed version of this kind of position has been worked out by van Fraassen (1980). Van Fraassen does not use the term "instrumentalist" to describe his view; he calls it "constructive empiricism." The term "constructive" is used by so many people that it often seems to have no meaning at all, so I have reserved it for the views discussed in section 12.5. I see van Fraassen's view as a version of the instrumentalist approach, but it does not matter much what we call it.

Van Fraassen suggests that all we should ask of theories is that they accurately describe the observable parts of the world. Theories that do this are "empirically adequate." An empirically adequate theory *might* also describe the hidden structure of reality, but whether or not it does so is of no interest to science. For van Fraassen, when a theory passes a lot of tests and becomes well established, the right attitude to have toward the theory is to "accept" it, in a special sense. To accept a theory is to (1) believe (provisionally) that the theory is empirically adequate, and to (2) use the concepts the theory provides when thinking about further problems and when trying to extend and refine the theory.

Regarding point 1, for a theory to be empirically adequate, it must describe *all* the observable phenomena that come within its domain, including those we have not yet investigated. Some of the familiar problems of induction and confirmation appear here. Regarding point 2, van Fraassen wants to recognize that scientists do come to "live inside" their theories; they make use of the theory's picture of the world when exploring new phenomena. Some versions of instrumentalism struggle to make sense of this fact. But van Fraassen says a scientist can "live inside" a theory while remaining agnostic about whether the theory is true.

How can we decide between van Fraassen's view and the version of scientific realism that I outlined earlier?

First we need to be sure that the two positions conflict. I said that *one* aim of science is to give us accurate representations of the world, including the unobservable parts. Van Fraassen says "science aims to give us theories which are empirically adequate" (1980, 12). So far, our views seem compatible. In some cases science could aim only at empirical adequacy, but in other cases it could aim at representing the hidden structure of the world as well.

And this is the *right* attitude for a realist to have. For various reasons and in various situations, it might make sense for a scientist to be cautious, or unconcerned, about the application of a theory to the unobserved structure of the world, even when he or she is becoming confident about empirical adequacy.

So van Fraassen has described an attitude that scientists can reasonably have toward *some* theories in *some* circumstances. But van Fraassen thinks that science *should aim at no more* than empirical adequacy.

As many have argued, one place where van Fraassen's view runs into trouble is the distinction between observable and unobservable parts of the world. Realists have argued that there is a continuum, rather than a sharp boundary, between the observable and the unobservable (Maxwell 1962). Some things can be observed with the naked eye, like trees. Other things, like the smallest subatomic particles, are unobservable and can only have their presence inferred from their effects on the behavior of observable things. But between the clear cases we have lots of unclear ones. Is it observation if you use a telescope? How about a light microscope? An X-ray machine? An MRI scan? An electron microscope? The realist thinks that the distinction between the observable and unobservable is vague, and not of the right kind to support general conclusions about what science aims to represent.

Van Fraassen accepts that the distinction between the observable and the unobservable is vague, and he accepts that there is nothing "unreal" about the unobservable. He also accepts that we learn about the boundary from science itself. Still, he argues, science is only concerned with empirical adequacy—making true claims about the observable part of the world. But this view cannot be defended. Van Fraassen is saying it's *never* reasonable for science to aim at describing the structure of the world beyond *this particular* boundary. Suppose we describe a slightly different boundary, based on a concept a bit broader than "observation." Let's say that something is *detectable* if either it is observable or its presence can be very reliably inferred from what is observable. As with van Fraassen's concept of observability, science itself tells us which things are detectable. In this sense, the chemical structures of various important molecules like sugars and

DNA are detectable although not observable. So why shouldn't science aim at giving us accurate representations of the detectable features of the world as well as the observable features? Why shouldn't science aim to tell us what the molecular structure of complex sugars is like?

Perhaps our beliefs about the detectable structures are not as reliable as our beliefs about the observable structures. If so, we need to be more cautious when we take theories to be telling us what the detectable structure of the world is like. But that is no problem; we often need to be cautious.

What is so special about the "detectable"? Nothing, of course. We could define an even broader category of objects and structures, which includes the detectable things plus those that can have their presence inferred from observations with *moderate* reliability. Why should science stop before trying to work out what lies beyond this boundary? We might need to be even *more* careful with our beliefs about those features of the world, but that is no problem.

You can see how the argument is going. From the realist point of view, there is no boundary that marks the distinction between features of the world that science can reasonably aim to tell us about and features that science cannot reasonably aim to tell us about. As we learn about the world, we also learn more and more about which parts of the world we can expect to have reliable information about. And there is no reason why science should not try to describe all the aspects of the world that we can hope to gain reliable information about. As we move from one area to another, we must often adjust our level of confidence. Sometimes, especially in areas such as theoretical physics, which are fraught with strange puzzles, we might have reason to adopt something like van Fraassen's attitude, at least temporarily. But it is a mistake to think that empirical adequacy of van Fraassen's kind is *the* aim of science.

OK

12.7 Representation, Models, and Truth (Optional Section)

I will finish the chapter with further discussion of an issue introduced in section 12.3. I formulated scientific realism by saying that science tries to give us "accurate representations" of the world. Most discussion of this topic in twentieth-century philosophy treated theories as *linguistic* entities, as collections of sentences. So when people tried to work out what sorts of relationships theories have with reality, they drew on concepts from the philosophy of language. In particular, the concepts of *truth* and *reference* were emphasized. A good scientific theory is a true theory; how can we determine which theories are true? Electrons exist if the word "electron" refers to them; how do we decide whether a term in a scientific theory refers to anything? A range of problems came to be addressed via the concepts of truth and reference.

This emphasis might be a bad idea. There are several issues to consider. One has to do with the "representational vehicles," or representational media, used by science. Science does express hypotheses about the world using sentences in language, either ordinary language or technical extensions of ordinary language. But in other cases, science uses representational vehicles of a different kind. Many hypotheses in science are expressed using *models.* Consider the case of mathematical models. These are abstract mathematical structures that are supposed to represent key features of real systems in the world. But in thinking about how a mathematical model might succeed in representing the world, the linguistic concepts of truth, falsity, reference, and so forth do not seem to be useful. Models have a different *kind* of representational relationship with the world from that found in language. A good model is one that has some kind of *similarity* relationship, probably of an abstract kind, with the system that the model is "targeted" at (Giere 1988). It is hard to work out the details of this idea.

The role of models in science did become an important topic in late-twentieth-century philosophy (Suppe 1977). Some argued that we should use the idea of a model to give a different description of how *all* theories work in science. But it is a mistake to think that all of science uses the same "vehicles" to represent the world. We should not replace a language-based analysis of all science with a model-based analysis. What we find in science is a range of different representational vehicles. OK

Consider Darwin's *Origin of Species.* Darwin's book contained a set of hypotheses about the world, supported with elaborate arguments, expressed using rather ordinary language. But not all science is like this. Even the topics that Darwin was addressing are now treated differently. Recent discussions of how natural selection changes biological populations tend to be expressed in the form of mathematical models. These models are written down, of course. They are formulated using mathematical symbolism, and they have to be supplemented with a commentary telling us (for example) which phenomena in the real world are being represented by the model. But we should not expect an analysis of how mathematical models relate to the world to use the same concepts as an analysis of how hypotheses expressed in ordinary language relate to the world.

Not all models in science are mathematical. More generally, we might think of a model as a structure that is intended to represent another structure by virtue of an abstract similarity relationship between them. Sometimes the aim might be to understand the unfamiliar by modeling it on the familiar (as in Bohr's early "solar system" model of the atom). But this is not always what is going on. Abstract mathematical models might be thought of as attempts to use a general-purpose and precise framework to represent *dependence relationships* that might exist between the parts of real systems. A mathematical model will treat one variable as a function of others, which in turn are functions of others, and so on. In this way, a complicated network of dependence structures can be represented. And then, via a commentary, the dependence structure in the model can be treated as representing the dependence structure that might exist in a real system.

Models, whether mathematical or not, have a kind of flexibility that is important in scientific work. A variety of people can use the same model while interpreting it differently. One person might use the model as a predictive device, something that gives an output when you plug in specific inputs, without caring how the inner workings of the model relate to the real world. Another person might treat the same model as a highly detailed picture of the dependence structure inside the real system being studied. And there is a range of possible attitudes between these two extremes; another person might treat the model as representing *some* features, but only a few, of what is going on in the real system.

The difference between models and linguistically expressed theories may be important in understanding progress in science. Many old scientific theories, now superseded, can look like failures when we ask whether much of the theory was *true,* and whether the terms in the theory *referred* to anything. But sometimes, if we recast the old theory as a model, we find that the model had some of the right *structure,* from the point of view of our current theories. Worrall (1989) uses the case of various "ether" theories from nineteenth-century physics; they had good structural features even though the ether does not exist.

In criticizing the emphasis on truth and reference in philosophy of science, I have stressed the role of representational vehicles that require a different kind of analysis. Some would add that even when we are dealing with language, the concepts of truth and reference might be bad ones to use.

Some philosophers think that to call a theory true is to assert that it has a special connection to the world. Traditionally, this has been described as a *correspondence* relationship. That term can be misleading, as it suggests a kind of "picturing," which is not what modern theories of truth propose. But this first option holds that there is some kind of special and valuable relationship between true theories and the world. If this is so, we can use the concept of truth when analyzing scientific language and its relations to reality. Others argue that the concept of truth is *not* suitable for this kind of use. The word "true" is one that we use to signal our agreement or disagreement with others, not to describe real connections between language

and the world (Horwich 1990). In sociology of science, Bloor (1999) has defended a position of this kind.

In this chapter I have been cautious about truth. I used a broad concept of "accurate representation" to describe a goal that science has for its theories. Some argue that even the idea of *representation* as a genuine relationship between symbols and the world is mistaken, whether the symbols are in language, models, thought, or whatever. That will sound like a radical position, and so it is. (This is one claim made by postmodernists, for example.) But it is hard to work out which theories about symbols retain the familiar idea of representation, and which do not.

Further Reading

Key works in the resurgence of scientific realism include Jack Smart's *Philosophy and Scientific Realism* (1963) and various papers collected in Hilary Putnam's *Mind, Language, and Reality* (1975). See also Maxwell 1962.

Leplin, *Scientific Realism* (1984), is a very good collection on the problem. Boyd's paper in that book is a useful survey of the options, with key differences from the one given here. Boyd also gives an influential defense of scientific realism. Devitt, *Realism and Truth* (1997), defends both common-sense and scientific realism. Psillos 1999 is a very detailed treatment of the debate.

For further discussion of the relations between realism and success, see Stanford 2000. On the problems raised by quantum physics, see Albert 1992. For a more detailed discussion of how avoidance of the "Bad View" has shaped sociology of science, see Godfrey-Smith 1996, chapter 5.

Churchland and Hooker, *Images of Science* (1985), is a good collection on van Fraassen.

Kitcher (1978) battles with the problems of meaning and reference for scientific language and their consequences for realism. See also Bishop and Stich 1998 on this problem. Lynch 2001 is a recent collection on the problem of truth.

There is a large literature on the role of models in science (Suppe 1977). Confusion sometimes arises because the usual sense of the word "model" in philosophy is different from that found in science itself (see the glossary). So different people wanting to "analyze science in terms of models" often have very different tasks in mind (Downes 1992). One useful and interesting treatment of the issue is in Giere's *Explaining Science* (1988, chapter 3). Hesse 1966 is a famous early discussion, focused, however, on yet another sense of "model."

Fine 1984 and Hacking 1983 are influential works on scientific realism that defend rather different views from those discussed here.

13

Explanation

13.1 Knowing Why

What does science do for us? In chapter 12 I argued for a version of scientific realism, according to which one aim of science is describing the real structure of the world. Science aims to tell us, and often succeeds in telling us, what the world is like. But it is also common to think that science tells us *why* things happen; we learn from science not just what goes on but why it does. Science apparently seeks to *explain* as well as describe. So we seem to face a new question. What *is* it for a scientific theory to explain something? In what sense does science give us an *understanding* of phenomena, as opposed to mere descriptions of what there is and what happens?

The idea that science aims at explanations of *why* things happen has sometimes aroused suspicion in philosophers, and it has also done so in scientists themselves. Such distrust is reasonably common within strong empiricist views. Empiricists have often seen science, most fundamentally, as a system of rules for predicting experience. When explanation is put forward as an *extra* goal for scientific theories, empiricists get nervous.

There is a complicated relationship between this problem of explanation and the problem of analyzing confirmation and evidence (chapters 3, 14). The hope has often been to treat these problems separately. Understanding evidence is problem 1; this is the problem of analyzing what it is to have evidence to believe that a scientific theory is true. Understanding explanation is problem 2; here we assume that we have already chosen our scientific theories, at least for now. We want to work out how our theories provide explanations. In principle, we can make a distinction of this kind. But there is a close connection between the issues. The solution to problem 2 may affect how we solve problem 1. Theories are often preferred by scientists because they seem to yield good explanations of puzzling phenomena. In chapter 3, explanatory inference was defined as inference from a set of data to a hypothesis about a structure or process that would explain the

data. This seems much more common in science than the traditional philosophical idea of inductive inference (inference from particular cases to generalizations). This suggests that there is a close relation between the problem of analyzing explanation and the problem of analyzing evidence.

There is a very large literature on explanation, but these issues will get a whirlwind treatment in this book. One reason for this is that I think the philosophy of science has approached the problem of explanation in a mistaken way. To some extent, that is true of many topics in this book; there have been plenty of wrong turns in the philosophy of science. But in the case of explanation, I think the error has been fairly clear; I will describe it in section 13.3. So some of the views presented in this chapter are rather unorthodox.

13.2 The Rise and Fall of the Covering Law Theory of Explanation

Empiricist philosophers, I said above, have sometimes been distrustful of the idea that science explains things. Logical positivism is an example. The idea of explanation was sometimes associated by the positivists with the idea of achieving deep metaphysical insight into the world—an idea they would have nothing to do with. But the logical positivists and logical empiricists did make peace with the idea that science explains. They did this by construing "explanation" in a low-key way that fitted into their empiricist picture.

The result was the *covering law* theory of explanation. This was the dominant philosophical theory about scientific explanation for a good part of the twentieth century. The view is now dead, but its rise and fall are instructive.

The covering law theory of explanation was first developed in detail by Carl Hempel and Paul Oppenheim in a paper (1948) that became a centerpiece of logical empiricist philosophy. Let us begin with some terminology. In talking about how explanation works, the *explanandum* is whatever is being explained. The *explanans* is the thing that is doing the explaining. If we ask "why X?" then X is the explanandum. If we answer "because Y," then Y is the explanans.

The basic ideas of the covering law theory are simple. Most fundamentally, to explain something is to *show how to derive it in a logical argument*. The explanandum will be the conclusion of the argument, and the premises are the explanans. A good explanation must first of all be a good logical argument, but in addition, the premises must contain at least one statement of a *law of nature*. The law must make a real contribution to the argument; it cannot be something merely tacked on. (Of course, for an explanation to be a good one in the fullest sense, the premises must also be *true*. But the first task here is to describe what sort of statements *would* give us a good explanation of a phenomenon, if the statements were true.)

Some explanations (both in science and in everyday life) are of particular events, while others are directed at general phenomena or regularities. For example, we might try to explain the particular fact that the U.S. stock market crashed in 1929, in terms of economic laws operating against the background conditions of the day. And we can also explain patterns; Newton is often seen as giving an explanation of Kepler's laws of planetary motion in terms of more basic laws of mechanics in combination with assumptions about the layout of the solar system. In both cases, the covering law theory sees these explanations as expressible in terms of arguments from premises to conclusions. Some of the arguments that express explanations will be deductively valid, but this is not required in all cases. The covering law theory was intended to allow that some good explanations could be expressed as nondeductive arguments ("inductive" arguments, in the logical empiricists' broad sense of the term). If we can take a particular phenomenon and embed it within an argument in which the premises include a law and bestow high probability on the conclusion, this yields a good explanation of the phenomenon.

There were many problems of detail encountered in attempts to formulate the covering law theory precisely (Salmon 1989). The problems were more difficult in the case of nondeductive arguments, and also in the case of explaining patterns rather than particular events. I won't worry about the technicalities here. The basic idea of the covering law theory is simple and clear: to explain something is to show how to derive it in a logical argument of a kind that makes use of a law in the premises. To explain something is to *show that it is to be expected,* to show that it is *not surprising,* given our knowledge of the laws of nature.

For the covering law theory, there is not much difference between explanation and prediction. To predict something, we put together an argument and try to show that it is to be expected, though we don't know for sure yet whether it is going to happen. When we explain something, we know that it has happened already, and we show that it *could* have been predicted, using an argument containing a law. You might be wondering at this point what a "law of nature" is supposed to be. This was a troubling topic for logical empiricism and has continued to be troubling for everyone else. But a "law of nature" was not supposed to be something very grandiose. It was supposed to be a kind of basic regularity, a basic pattern, in the flow of events. (I return to this question in section 13.4.)

Though I use the phrase "covering law theory" here, another name for the theory is the "D-N theory," or "D-N model," of explanation. "D-N" stands for "deductive-nomological," where the word "nomological" is from the Greek word for law, *nomos*. The term "D-N" can be confusing because, as I said, the argument in a good explanation need not be deductive. So "D-N" really only refers to some covering law explanations, the deductive ones.

That concludes my sketch of the covering law theory. I now move on to what is wrong with it. This is a case where we have something close to a knockdown argument. Although there are many famous problems with the covering law theory, the most convincing problem is usually called the *asymmetry problem*. And the most famous illustration of the asymmetry problem is the case of the flagpole and the shadow.

Suppose we have a flagpole casting a shadow on a sunny day. Someone asks: why is the shadow *X* meters long? According to the covering law theory, we can give a good explanation of the shadow by deducing the length of the shadow from the height of the flagpole, the position of the sun, the laws of optics, and some basic trigonometry. We can show why that length of shadow was to be expected, given the laws and the circumstances. The argument can even be made deductively valid. So far, so good. The problem is that we can run just as good an argument in another direction. Just as we can deduce the length of the shadow from the height of the pole (plus optics and trigonometry), *we can deduce the height of the pole from the length of the shadow* (and the same laws). An equally good *argument,* logically speaking, can be run in both directions; either can give information about the other. But it seems that we cannot run an equally good *explanation* in both directions, though the covering law theory says we can. It is fine to explain the length of the shadow in terms of the flagpole and the sun, but it is not fine to explain the length of the flagpole in terms of the shadow and the sun. (At least, it is not fine unless this is a very unusual flagpole—perhaps one that is designed to regulate its own length to maintain a particular shadow.)

What we find here is that explanations have a kind of directionality. Some arguments (though not all) can be reversed and remain good as arguments. But explanations cannot be reversed in this way (except in some special cases). So not all good arguments that contain laws are good explanations. This objection to the covering law theory was famously given (using a slightly different example) by Sylvain Bromberger (1966).

Once this point is seen, it becomes obvious and devastating. The covering law theory sees explanation as very similar to prediction; the only difference is what you know and what you don't know. But this is a mistake. Consider the concept of a *symptom*. Symptoms can be used to predict, but they cannot be used to explain. Yet a symptom can often be used in a good logical argument, along with a law, to show that something is to be expected.

If you know that only disease D produces symptom S, then you can make inferences from S to D. You might in some cases be able to make predictions from D to S, too. But you cannot *explain* a disease in terms of a symptom. Explanation only runs one way, from D to S, no matter how many different kinds of inferences can be made in other directions. And, further, it seems that good explanations of S can be given in terms of D even if S is *not* a very reliable symptom of D, even if S is not always to be expected when someone has D. This is a separate problem for the covering law theory, often discussed using the example of some unreliable but unpleasant symptoms of syphilis.

In some of these cases, the covering law theory can engage in fancy footwork to evade the problem. But other cases, including the original flagpole case, seem immune to footwork. Hempel's own attitude to the issue was puzzling. He actually anticipated the problem, but dismissed it (Hempel 1965, 352–54). His strategy was to accept that if his theory allowed explanations to run in two directions in cases where it seems that explanation only runs in one direction, then both directions must really be OK. In some actual scientific cases this reply seems reasonable; there are cases in physics where it *is* hard to tell which direction(s) the explanation(s) are running in. But in other cases the direction seems completely clear. In the case of the flagpole and the shadow, this reply seems hopeless.

There are other good arguments against the covering law theory (Salmon 1989), but the asymmetry problem is the killer. It also seems to be pointing us toward a better account of explanation.

13.3 Causation, Unification, and More

What is it about the flagpole's height that makes it a good explanation for the length of the shadow, and not vice versa? The answer seems straightforward. The shadow is *caused* by the interaction between sunlight and the flagpole. That is the direction of causation in this case, and that is the direction of explanation also. So we seem to get an immediate suggestion from the flagpole case for how to build a better theory: to explain something is to describe what caused it. Why did the dinosaurs become extinct 65 million years ago? Here again, our request for an explanation seems to amount to a request for information about what caused the extinction.

Although that conclusion seems compelling, it has not been universally accepted, and it raises many further problems. The biggest question raised at this point is, What *is* causation? We confidently used the idea of causation to resolve the flagpole case, but the whole idea of causation and causal connection is extremely controversial in philosophy. For many philosophers, causation is a suspicious metaphysical concept that we do best to avoid when trying to understand science. This suspicion is, again, common within the empiricist tradition. It derives from the work of Hume. The suspicion is directed especially at the idea of causation as a sort of hidden connection between things, unobservable but essential to the operation of the universe. Empiricists have often tried to understand science without supposing that science concerns itself with alleged hidden connections of that kind. The rise of scientific realism in the latter part of the twentieth century led to some easing of this anxiety. But many philosophers would be pleased to see an adequate account of science that did not get entangled with issues about causation.

Despite this unease, toward the end of the twentieth century, the main proposal about explanation being discussed, in different ways, was the idea that explaining something is giving information about how it was caused. Some sophisticated analyses were developed that sought to use probability theory to clarify this basic idea (Salmon 1984; Suppes 1984). It might seem initially that this view of explanation is most directly applied to explanations of particular events (like the extinction of the dinosaurs), but it can also be applied to the explanation of patterns. We can ask, Why does inbreeding produce an increase in birth defects? The explanation will describe a general kind of causal process that is involved in producing the phenomenon (a process involving an increased chance that two copies of a recessive gene will be brought together in a single individual).

Claiming that causation is the key to explanation does not settle all the issues about explanation. We need to know what *kind* of information about causes is needed for a good explanation. One way to think of the situation is to imagine an idealized "complete" explanation that contains *everything* in the causal history of the event to be explained, specified in total detail (Railton 1981). No one ever wants to be told the complete explanation for a phenomenon, and we never know these complete explanations. Instead, in any context of discussion in which a request for an explanation is made, some *piece* or *pieces* of the complete explanation will be relevant. We are often able to know, and describe, these relevant pieces of a total causal structure. To give a good explanation in actual practice, all that is required is a description of these relevant pieces of the whole.

One main alternative to the analysis of explanation in terms of causation was developed in the years after the demise of the covering law theory. This was the idea that explanation should be analyzed in terms of *unification*. This idea was developed in detail by Michael Friedman (1974) and Philip Kitcher (1981, 1989). But as Kitcher also emphasized, the idea was actually present all along in logical empiricism. The idea that explanation is unification was a sort of "unofficial" theory of explanation within much logical empiricism, in contrast to the "official" covering law theory (see, for example, Feigl 1943). This unofficial theory is a good deal better than the official theory. Often the two approaches were mixed in together; to show the connection between particular events and a general law is, after all, to achieve a kind of unification. Why not develop a theory of unification in science that is not tied to the idea of deriving phenomena from laws?

So the *unificationist* theory holds that explanation in science is a matter of connecting a diverse set of facts by subsuming them under a set of basic patterns and principles. Science constantly strives to reduce the number of things that we must accept as fundamental. We try to develop general *explanatory schemata* (explanatory schemes) that can be applied as widely as possible. This proposal certainly makes a lot of sense of how scientists operate. Indeed, it seems clear that what produces an "Aha!" reaction is often the realization that some odd-looking phenomenon is really a case of something more general. Kitcher also argues for this view with cases from the history of science. He argues that some very famous theories—Darwin's theory of evolution and Newton's later work on the nature of matter—were compelling to scientists in their early stages despite not making many specific new predictions, because they promised to *explain so much*. And this "explanatory promise" seems to have been the ability of those theories to *unify* a great range of phenomena with a few general principles.

In Kitcher's case, another reason for developing the unificationist theory was a distrust of the idea of causation. This led Kitcher to try, for some years, to develop a theory of explanation *entirely* in terms of unification. But what about the flagpole and the shadow, and the asymmetry in which can explain which? Kitcher argued that we do tend to describe this asymmetry in causal terms, but this causal talk is really a loose summary of more basic asymmetries that involve unification (Kitcher 1989).

So we have two main proposals to replace the covering law theory: the causal theory and the unification theory. These have often been treated as competitors: "Does causation win or does unification win?" But this is surely a mistake. We do not have to choose. Again, beware the dubious allure of simplicity in philosophical theories! Much of the time, to explain something is to describe the causal mechanisms behind it or the causal history leading up to it. That is true *much* of the time, but there is no need to hold that it is true *all* the time. In some cases there can be pretty clear explanatory relations between patterns or principles, even when causal language is hard to apply to the situation. Often this seems to involve unification. Nothing stops us from holding that a variety of different relations can be explanatory.

Recently, ideas similar to this have been emerging in the philosophy of science. Wesley Salmon was for many years one of the main partisans for the idea that causation is the key to explanation. But he eventually accepted that unification is also part of the story. Sometimes he seemed to think of causation and unification as two sides of the same explanatory coin, and some other times as alternative explanatory projects (1989, 1998). Kitcher, who tried for years to avoid using the idea of causation to analyze explanation, instead telling the whole story in terms of unification, has now decided that this was probably a mistake and the concept of causation is not so dubious after all (personal communication, 2002).

So what might be emerging is a kind of "pluralism" about explanation in the philosophy of science. This is a step in the right direction, but I suggest that the whole issue has been approached wrongly. (This is where I become unorthodox.) The most peculiar thing about the discussion of explanation by philosophers has been the assumption that explanation is the kind of thing that requires analysis in terms of a single special relation or a short list of special relations. It is a mistake to think there is one basic relation that is *the* explanatory relation (as in the covering law theory, the causal theory, and the unification theory), and it is also a mistake to think that there are some definite two or three such relations.

The alternative view is to recognize that the idea of explanation operates differently within different parts of science—and differently within the same part of science at different times. The word "explanation" is used in science for something that is sought, and sometimes achieved, by the development of theories, but what exactly is being sought is not constant in all of science. And we cannot get the right analysis by claiming that within all of science, a good explanation is something that satisfies *either* the causal test *or* the unification test (etc.). This familiar form of "pluralism" leaves out the way that different scientific fields will establish definite criteria for what will pass as a good explanation. The standards for a good explanation in field A need not suffice in field B. If an "ism" is required, the right analysis of explanation is a kind of *contextualism*—a view that treats the standards for good explanation as partially dependent on the scientific context.

Kuhn argued some years ago for a view of this kind (1977a). In a paper about the history of physics, he claimed that different theories (or paradigms) tend to bring with them their own standards for what counts as a good explanation. He argued, further, that standards about whether a relation counts as "causal" also depend on paradigms. The concepts of explanation and causation are, to some extent at least, internal to different scientific fields and historical periods.

In the case of causation, a philosopher might reply to Kuhn, with some justification, that just because different people have *thought* differently about what causation is does not mean that there *is* no fact of the matter. Fair enough (at least for now). But in the case of explanation, I think this reply has little force. If two scientific fields single out different relations and call them "explanation," there need be no factual error that one or the other is making.

To support this claim, Kuhn focused (as he did in *Structure*) on the case of Newton's theory of gravity. Does Newton's theory *explain* the falling of objects, given that Newton's treatment of gravity gave no intuitive mechanism but only a mathematical relationship? Some answered no, but over time it became part of Newtonianism that the right kind of mathematical law *does count* as an explanation. It is Kuhn's view that the idea of explanation will evolve as our ideas about science and about the universe change.

So although the covering law theory definitely fails as a general account of explanation in science, it would be a mistake to conclude that *no* explanations have the form described by the covering law theory. There are some explanations that are at least close to what Hempel had in mind. The mistake is to apply that model to all cases.

I suggest that Kuhn was right on this point. I add that this proposal need not lead to the radical idea that *anything* can count as an explanation. Scientific traditions will generally have good reasons for their treatment of the idea of explanation; views about explanation will depend on views about what the world contains, for example. Some possible concepts of explanation will embed factual errors. To use a simple example, if someone claims that good explanations are always based on a concept of God's will, but it turns out that there is no God, then that conception of explanation will be mistaken because of a factual error. Some philosophers might make the same argument about concepts of explanation that use the idea of causation—they might argue that the traditional idea of causal connection is a piece of mistaken metaphysics. But many possible treatments of the idea of explanation will not be ruled out by factual errors.

This is a case where it is important to pay attention to the actual use of the term "explain" in science. Here we find a lot of diversity. In some fields, there are technical senses of the word, even mathematical measurements of "the amount of variation explained" by a given factor. In other fields, nothing like a technical standard applies. The word "explain" also has an almost rhetorical use. Someone might say: "your theory does accommodate this result, but it does not really explain it." This might mean "your theory can only be used to derive this result in an unnatural-looking way." (Often, unification seems important in cases like this.) At other times the word "explain" is used in a much more low-key way in science. According to scientific realism, a lot of science is aimed at describing what is going on in the world; often this will be a matter of describing *how things work*. How does photosynthesis work? How does the replication of DNA work? Descriptions of phenomena of this kind will often be referred to as explanations, but this does not mean that something *extra* is going on, beyond the description of mechanisms and processes.

At this point I should compare my view with another unorthodox position in this area, that of van Fraassen (1980). He denies, as I do, that explanation is some single, special relation common to all of science. He has developed a "pragmatic" account of explanation, in which what counts as an explanation varies according to context. But this is very different from my view. Van Fraassen wants to deny that explanation is something "inside" science at all; he denies that scientific reasoning includes the assessment of the explanatory power of theories. Instead, explanation is something that people do when they take scientific theories and use them to answer questions that are external to scientific discussion itself. In contrast, the view I am defending is a view in which explanation is thoroughly internal to science, *but variously so*. Assessments of what explains what are a very important part of scientific reasoning, but different fields may use somewhat different concepts and standards of explanation.

Before leaving this topic, I should also add a comment about *explanatory inference*. Back in chapter 3, I used this term for inference from a set of data to a hypothesis about a structure or process that would explain the data. In chapter 14, when I look at more recent discussions of confirmation and evidence, I will return to this topic. The term "explanatory inference" suggests that there is one kind of relationship between data and hypothesis—*the* explanation relationship—that is involved in explanatory inference. Many philosophers would accept this. The term "inference to the best explanation" is, in fact, a more common name for what I call explanatory inference; that term suggests that there is some single measure of "explanatory goodness" involved. But I think this is the wrong way to think about scientific reasoning (and this is why I have avoided the term "inference to the best explanation"). I use "explanatory inference" in a broader way that does not suppose that there is a single measure of explanatory goodness involved, which applies to all of science. Rather, explanatory inference is a matter of devising and comparing hypotheses about hidden structures that might be responsible for data. "Explanation" is seen as something pretty diverse.

To sum up: the covering law theory is dead, as a general account of explanation in science. But we should not look for a new theory of some single relation or pattern that is involved in all scientific explanation. Very often, causation is involved. The same goes for unification and for deriving phenomena from laws. But different fields have different concepts and standards of explanation.

13.4 Laws and Causes (Optional Section)

This short and rather abstract section is a digression from the main themes of the book, a foray into an intersection between philosophy of science and the controversial field of *metaphysics*. The covering law theory of explanation made use of the idea of a law of nature. One of the theories that replaced it made use of the idea of causation. But what are laws of nature? What is causation?

In both cases, we have concepts that seem aimed at picking out a special kind of *connection* between things in the world. Causation is sometimes called, half jokingly, "the cement of the universe" (Mackie 1980; the phrase was first used by Hume [(1740) 1978]). In recent years, many philosophers have been skeptical about these concepts. But generally, their attitude has not been to reject them ("There is no such thing as causation") but instead to reconstrue these concepts in a very low-key way ("Yes, there is causation, but it is no more than this . . ."). In particular, philosophers have tried to analyze both laws and causation in terms of patterns in the *arrangement* of things, rather than some extra connection *between* things. Sometimes this project is referred to as "Humeanism," after David Hume, the first philosopher to develop a really focused suspicion about concepts of connection between events in nature (see also Lewis 1986b). The present-day Humeans do not have the same kind of empiricism as Hume, but they do want to avoid believing in any sort of unobservable cement connecting the universe together.

So a philosopher with Humean views will try to construe laws of nature as no more than regularities, or basic patterns, in the arrangement of events. To treat laws of nature in this way is to leave behind one of the familiar connotations of the term "law." Usually, we see laws as directing, or guiding, or governing in some way. It is possible (indeed traditional) to see laws of nature as governing the flow of events in the universe. Laws are seen as *responsible* for the regular patterns that we see, rather than being *identical* with those patterns (Dretske 1977; Armstrong 1983). The Humean regards this "governing" conception of laws as a seduction that must be avoided by hard-headed philosophers. The logical empiricist attitude toward laws of nature was basically Humean, in this sense.

The topic of causation has generated a similar debate. On one side we have those who see causation as basically some special kind of regular pattern in the flow of events. One the other side are those who see causation as a connection between things that is somehow *responsible* for the patterns (see Sosa and Tooley 1993). Perhaps this connection need not be seen as a mysterious philosophical entity; maybe it can be described by ordinary science (Dowe 1992; Menzies 1996).

For some years philosophers tended to discuss laws and their role in science in a way that had little contact with actual scientific work. In 1983 Nancy Cartwright delivered a wake-up call to the field with a book called *How the Laws of Physics Lie,* in which she argued that what people call "laws of physics" do not usually describe the behavior of real systems at all, but only describe the behavior of highly idealized fictional systems. Another important change that resulted from a closer look at actual science is that philosophers are no longer obsessed with natural laws as *the* goal of scientific theorizing. Over many years philosophers searched fields like biology for statements of laws of nature. Philosophers thought that any genuine science had to contain hypothesized laws and had to organize its ideas via the concept of a law. In fact, most biology has little use for the concept of a law of nature, but that does not make it any less scientific.

Further Reading

Wesley Salmon's *Four Decades of Scientific Explanation* (1989) is a very good survey of work on explanation between 1948 (the advent of the covering law theory) and the late 1980s. (The only thing marring Salmon's discussion is his rather eccentric theory about causation, which affects his treatment of explanation.)

A very good alternative discussion of causation and explanation can be found in Lewis 1986a. (Lewis's theory of causation is also eccentric. In fact, I guess *every* philosopher's theory of causation is eccentric; no two philosophers seem to agree. Lewis's discussion is compatible with a range of different views about causation, though.)

For unificationist theories of explanation, see Friedman 1974 and Kitcher 1989. These are fairly advanced papers.

Lewis discusses the Humean program in metaphysics in the preface to his 1986 *Philosophical Papers,* volume 2. Armstrong 1983 is a clear introduction to the more purely philosophical side of the literature on laws of nature. Beebee 2000 is a good discussion of the idea that laws "govern" things. Mitchell (2000) defends an interesting position on laws.