3

Induction and Confirmation

3.1 The Mother of All Problems

In this chapter we begin looking at a very important and difficult problem, the problem of understanding how observations can *confirm* a scientific theory. What connection between an observation and a theory makes that observation *evidence for* the theory? In some ways, this has been *the* fundamental problem in the last hundred years of philosophy of science. This problem was central to the projects of logical positivism and logical empiricism, and it was a source of constant frustration for them. And although some might be tempted to think so, this problem does not disappear once we give up on logical empiricism. The problem, in some form or other, arises for nearly everyone.

The aim of the logical empiricists was to develop a *logical* theory of evidence and confirmation, a theory treating confirmation as an abstract relation between sentences. It has become fairly clear that their approach to the problem is doomed. The way to analyze testing and evidence in science is to develop a different kind of theory. But it will take a lot of discussion, in this and later chapters, before the differences between approaches that will and will not work in this area can emerge. The present chapter will mostly look at how the problem of confirmation was tackled in the middle of the twentieth century. And that is a tale of woe.

Before looking at twentieth-century work on these issues, we must again look further into the past. The confirmation of theories is closely connected to another classic issue in philosophy: *the problem of induction*. What reason do we have for expecting patterns observed in our past experience to hold also in the future? What justification do we have for using past observations as a basis for generalization about things we have not yet observed?

The most famous discussions of induction were written by the eighteenth-century Scottish empiricist David Hume ([1739] 1978). Hume asked, What reason do we have for thinking that the future will resemble the past? There is no *contradiction* in supposing that the future could be totally unlike the past. It is *possible* that the world could change radically at any point, rendering previous experience useless. How do we know this will not happen? We might say to Hume that when we have relied on past experience before, this has turned out well for us. But Hume replies that this is begging the question—presupposing what has to be shown. Induction has worked in the past, sure, but that's the *past*! We have successfully used "past pasts" to tell us about "past futures." But our problem is whether *anything* about the past gives us good information about what will happen *tomorrow.*

Hume concluded that we have no reason to expect the past to resemble the future. Hume was an "inductive skeptic." He accepted that we all use induction to make our way around the world. And he was not suggesting that we stop doing so (even if we could). Induction is psychologically natural to us. Despite this, Hume thought it had no rational basis. Hume's inductive skepticism has haunted empiricism ever since. The problem of confirmation is not the same as the classical problem of induction, but it is closely related.

3.2 Induction, Deduction, Confirmation, and Explanatory Inference

The logical empiricists tried to show how observational evidence could provide support for a scientific theory. The idea of "support" is important here; there was no attempt to show that scientific theories could be *proved*. Error is always possible, but evidence can support one theory over another.

The cases that were to be covered by this analysis included the simplest and most traditional cases of induction: if we see a multitude of cases of white swans, and no other colors, why does that give us reason to believe that all swans are white? But obviously not all cases of evidence in science are like this. The observational support for Copernicus's theory that the earth goes around the sun, or for Darwin's theory of evolution, seems to work very differently. Darwin did *not* observe a set of individual cases of evolution and then generalize.

The logical empiricists wanted a theory of evidence, or "theory of confirmation," that would cover all these cases. They were not trying to develop a *recipe* for confirming theories. Rather, the aim was to give an account of the relationships between the statements that make up a scientific theory and statements describing observations, which make the observations support the theory. You might wonder, at this point, what use there could be for a theory with so distant a relationship to actual scientific behavior. Who cares whether a logical analysis of this kind exists or not? In defense of logical empiricism, we might say this: although scientific behavior is not being directly described by the theory of confirmation, nonetheless scientific procedures might be *based* on assumptions described in the theory of confirmation. Perhaps scientists do many things that cannot be justified if confirmation does not exist.

Let us look more closely at what the logical empiricists tried to do. First, I should say more about the distinction between deductive and inductive logic (a distinction introduced in chapter 2). Deductive logic is the well-understood and less controversial kind of logic. It is a theory of patterns of argument that transmit truth with certainty. These arguments have the feature that *if* the premises of the argument are true, the conclusion is guaranteed to be true. An argument of this kind is *deductively valid.* The most famous example of a logical argument is a deductively valid argument:

PREMISES	All men are mortal.
	Socrates is a man.
	———————
CONCLUSION	Socrates is mortal.

A deductively valid argument might have false premises. In that case the conclusion might be false as well (although it also might not be). What you get out of a deductive argument depends on what you put in.

The logical empiricists loved deductive logic, but they realized that it could not serve as a complete analysis of evidence and argument in science. Scientific theories do have to be logically *consistent,* but this is not the whole story. Many inferences in science are not deductively valid and give no guarantee. But they still can be *good* inferences; they can still provide *support* for their conclusions.

For the logical empiricists, there is a reason why so much inference in science is not deductive. As empiricists, they believed that all our evidence derives from observation. Observations are always of *particular* objects and occurrences. But the logical empiricists thought that the great aim of science is to discover and establish *generalizations.* Sometimes the aim was seen as describing "laws of nature," but this concept was also regarded with some suspicion. The key idea was that science aims at formulating and testing generalizations, and these generalizations were seen as having an infinite range of application. No finite number of observations can conclusively establish a generalization of this kind, so these inferences from observations in support of generalizations are always nondeductive. (In contrast, all it takes is *one* case of the right kind to prove a generalization to be *false;* this fact will loom large in the next chapter.)

In many discussions of these topics, the logical empiricists (and some later writers) used a simple terminology in which all arguments are either deductive or inductive. Inductive logic was thought of as a theory of *all* good arguments that are not deductive. Carnap, especially, used "induction" in a very broad way. But this terminology can be misleading, and I will set things up differently.

I will use the term "induction" only for inferences from particular observations in support of generalizations. To use the most traditional example, the observation of a large number of white swans (and no swans of any other color) might be used to support the hypothesis that all swans are white. We could express the premises with a list of particular cases—"Swan 1 observed at time t_1 was white; swan 2 observed at time t_2 was white. . . ." Or we might simply say: "All the many swans observed so far have been white." The conclusion will be the claim that all swans are white—a conclusion that could well be false but which is supported, to some extent, by the evidence. Sometimes "enumerative induction" or "simple induction" is used for inductive arguments of this most traditional and familiar kind. Not all inferences from observations to generalizations have this very simple form, though. (And a note to mathematicians: *mathematical induction* is really a kind of *de*duction, even though it has the superficial form of *in*duction.)

A form of inference closely related to induction is *projection.* In a projection, we infer from a number of observed cases to arrive at a prediction about the *next* case, not to a generalization about all cases. So we see a number of white swans and infer that the next swan will be white. Obviously there is a close relationship between induction and projection, but (surprisingly, perhaps) there are a variety of ways of understanding this relationship.

Clearly there are other kinds of nondeductive inference in science and everyday life. For example, during the 1980s Luis and Walter Alvarez began claiming that a huge meteor had hit the earth about 65 million years ago, causing a massive explosion and dramatic weather changes that coincided with the extinction of the dinosaurs (Alvarez et al. 1980). The Alvarez team claimed that the meteor caused the extinctions, but let's leave that aside here. Consider just the hypothesis that a huge meteor hit the earth 65 million years ago. A key piece of evidence for this hypothesis is the presence of unusually high levels of some rare chemical elements, such as iridium, in layers in the earth's crust that are about 65 million years old. These chemical elements tend to be found in meteors in much higher concentrations than they are near the surface of the earth. This observation is taken to be strong evidence supporting the Alvarez theory that a meteor hit the earth around that time.

If we set this case up as an argument, with premises and a conclusion, it clearly is not an induction or a projection. We are not inferring to a generalization, but to a hypothesis about a structure or process that would explain the data. A variety of terms are used in philosophy for inferences of this kind. C. S. Peirce called these "abductive" inferences as opposed to inductive ones. Others have called them "explanatory inductions," "theoretical inductions," or "theoretical inferences." More recently, many philosophers have used the term "inference to the best explanation" (Harman 1965; Lipton 1991). I will use a slightly different term—"explanatory inference."

So I will recognize two main kinds of nondeductive inference, *induction* and *explanatory inference* (plus *projection,* which is closely linked to induction). The problem of analyzing confirmation, or the problem of analyzing evidence, includes all of these.

How are these kinds of inference related to each other? For logical positivism and logical empiricism, induction is the most fundamental kind of nondeductive inference. Reichenbach claimed that all nondeductive inference in science can be reconstructed in a way that depends only on a form of inference that is close to traditional induction. What looks like an explanatory inference can be somehow broken down and reconstructed as a complicated network of inductions and deductions. Carnap did not make this strong claim, but he did seem to view induction as a *model* for all other kinds of nondeductive inference. Understanding induction was in some sense the key to the whole problem. And the majority of the logical empiricist literature on these topics was focused on induction rather than explanatory inference.

So one way to view the situation is to see induction as fundamental. But it is also possible to do the opposite, to claim that explanatory inference is fundamental. Gilbert Harman argued in 1965 that inductions are justified only when they are explanatory inferences in disguise, and others have followed up this idea in various ways.

Explanatory inference seems much more common than induction within actual science. In fact, you might be wondering whether science contains *any* inductions of the simple, traditional kind. That suspicion is reasonable, but it might go too far. Science does contain inferences that look like traditional inductions, at least on the face of them. Here is one example. During the work that led to the discovery of the structure of DNA by James Watson and Francis Crick, a key piece of evidence was provided by "Chargaff's rules." These "rules," described by Erwin Chargaff in 1947, have to do with the relation between the amounts of the four "bases," C, A, T, and G, that help make up DNA. Chargaff found that in the DNA samples he analyzed, the amounts of C and G were always roughly the same, and the amounts of T and A were always roughly the same. This fact about DNA became important in the discussions of how DNA molecules are put together. I called it a "fact" just above, but of course Chargaff in 1947 had not observed all the molecules of DNA that exist, and neither have we. In 1947 Chargaff's claim rested on an induction from a small number of cases (in just eight different kinds of organisms). Today we can give an argument for why Chargaff's rules hold that is not just a simple induction; the structure of DNA explains why Chargaff's rules must hold. But it might appear that, back when the rules were originally discovered, the only reason to take the rules to describe all DNA was inductive.

So it might be a good idea to refuse to treat one of these kinds of inference as "more fundamental" than the other. Maybe there is more than one kind of good nondeductive inference (and perhaps there are others besides the ones I have mentioned). Philosophers often find it attractive to think that there is ultimately just one kind of nondeductive inference, because that seems to be a simpler situation. But the argument from simplicity is unconvincing.

Let us return to our discussion of how the problem was handled by the logical empiricists. They used two main approaches. One was to formulate an inductive logic that looked as much as possible like deductive logic, borrowing ideas from deductive logic whenever possible. That was Carl Hempel's approach. The other approach, used by Rudolf Carnap, was to apply the mathematical theory of probability. In the next two sections of this chapter, I will discuss some famous problems for logical empiricist theories of confirmation. The problems are especially easy to discuss in the context of Hempel's approach, which was simpler than Carnap's. A detailed examination of Carnap is beyond the scope of this book. Through his career, Carnap developed very sophisticated models of confirmation using probability theory applied to artificial languages. Problems kept arising. More and more special assumptions were needed to make the results come out right. There was never a knockdown argument against him, but the project came to seem less and less relevant to real science, and it eventually ran out of steam (Howson and Urbach 1993).

Although Carnap's approach to analyzing confirmation did not work out, the idea of using probability theory to understand confirmation remains popular and has been developed in new ways. Certainly this looks like a good approach; it does seem that observing the raised iridium level in the earth's crust made the Alvarez meteor hypothesis *more probable* than before. In chapter 14 I will describe new ways to use probability theory to understand the confirmation of theories.

Before moving on to some famous puzzles, I will discuss a simple proposal that may have occurred to you.

The term *hypothetico-deductivism* is used in several ways by people writing about science. Sometimes it is used to describe a simple view about testing and confirmation. According to this view, hypotheses in science are confirmed when their logical consequences turn out to be true. This idea covers a variety of cases; the confirmation of a white-swan generalization by observing white swans is one case, and another is the confirmation of a hypothesis about an asteroid impact by observations of the true consequences of this hypothesis.

As Clark Glymour has emphasized (1980), an interesting thing about this idea is that it is hopeless when expressed in a simple way, but something like it seems to fit well with many episodes in the history of science. One problem is that a scientific hypothesis will only have consequences of a testable kind when it is combined with other assumptions, as we have seen. But put that problem aside for a moment. The suggestion above is that a theory is confirmed when a true statement about observables can be derived from it. This claim is vulnerable to many objections. For example, any theory *T* deductively implies *T-or-S*, where *S* is any sentence at all. But *T-or-S* can be conclusively established by observing the truth of *S*. Suppose *S* is observational. Then we can establish *T-or-S* by observation, and that confirms *T*. This is obviously absurd. Similarly, if theory *T* implies observation *E*, then the theory *T&S* implies *E* as well. So *T&S* is confirmed by *E*, and *S* here could be anything at all. (Note the similarity here to a problem discussed at the beginning of section 2.4.) There are many more cases like this.

The situation is strange, and some readers may feel exasperation at this point. People do often regard a scientific hypothesis as supported when its consequences turn out to be true; this is taken to be a routine and reasonable part of science. But when we try to summarize this idea using simple logic, it seems to fall apart. Does the fault lie with the original idea, with our summary of the idea using basic logic, or with basic logic itself? The logical empiricist response was to hang steadfastly onto the logic, and often to hang onto their translations of ideas about science into a logical framework as well. This led them to question or modify some very reasonable-looking ideas about evidence and testing. But it is hard to work out where the fault really lies.

A related feature of logical empiricism is the use of simplified and artificial cases rather than cases from real science. The logical empiricists sought to strip the problem of confirmation down to its bare essentials, and they saw these essentials in formal logic. But to many, philosophy of science seemed to be turning into an exercise in "logic-chopping" for its own sake. And as we will see in the next sections, even the logic-chopping did not go well.

Despite this, there is a lot to learn from the problems faced by logical empiricism. Confirmation really *is* a puzzling thing. Let us look at some famous puzzles.

3.3 The Ravens Problem

The logical empiricists put much work into analyzing the confirmation of generalizations by observations of their instances. At this point we will switch birds, in accordance with tradition. How is it that repeated observations of black ravens can confirm the generalization that all ravens are black?

First I will deal with a simple suggestion that will not work. Some readers might be thinking that if we observe a large number of black ravens and no nonblack ones, then at least we are cutting down the number of ways in which the hypothesis that all ravens are black might be wrong. As we see each raven, there is one less raven that might fail to fit the theory. So in some sense, the chance that the hypothesis is true should be slowly increasing. But this does not help much. First, the logical empiricists were concerned to deal with the case where generalizations cover an infinite number of instances. In that case, as we see each raven we are not reducing the number of ways in which the hypothesis might fail. Also, note that even if we forget this problem and consider a generalization covering just a finite number of cases, the kind of support that is analyzed here is a very weak one. That is clear from the fact that we get no help with the problem of *projection.* As we see each raven we know there is one less way for the generalization to be false, but this does not tell us anything about what to expect with the *next* raven we see.

So let us look at the problem differently. Hempel suggested that, as a matter of logic, all observations of black ravens confirm the generalization that all ravens are black. More generally, any observation of an F that is also G supports the generalization "All F's are G." He saw this as a basic fact about the logic of support.

This looks like a reasonable place to start. And here is another obvious-looking point: any evidence that confirms a hypothesis *H* also confirms any hypothesis that is logically equivalent to *H*.

What is logical equivalence? Think of it as what we have when two sentences say the same thing in different terms. More precisely, if *H* is logically equivalent to H^*, then it is impossible for H to be true but H^* false, or vice versa.

But these two innocent-looking claims generate a problem. In basic logic the hypothesis "All ravens are black" is logically equivalent to "All nonblack things are not ravens." Let us look at this new generalization. "All nonblack things are not ravens" seems to be confirmed by the observation of a white shoe. The shoe is not black, and it's not a raven, so it fits the hypothesis. But given the logical equivalence of the two hypotheses, anything that confirms one confirms the other. So the observation of a white shoe confirms the hypothesis that all ravens are black! That sounds ridiculous. As Nelson Goodman (1955) put it, we seem to have the chance to do a lot of "indoor ornithology"; we can investigate the color of ravens without ever going outside to look at one.

This simple-looking problem is hard to solve. Debate about it continues. Hempel himself was well aware of this problem—he is the one who originally thought of it. But there has not been a solution proposed that everyone (or even most people) have agreed upon.

One possible reaction is to accept the conclusion. This was Hempel's response. Observing a white shoe *does* confirm the hypothesis that all ravens are black, though presumably only by a tiny amount. Then we can keep our simple rule that whenever we have an "All F's are G" hypothesis, any observation of an F that is G confirms it and also confirms everything logically equivalent to "All F's are G." Hempel stressed that, logically speaking, an "All F's are G" statement is not a statement about F's but a statement about everything in the universe—the statement that if something is an F then it is G. We should note that according to this reply, the observation of the white shoe also confirms the hypothesis that all ravens are green, that all aardvarks are blue, and so on. Hempel was comfortable with this situation, but most others have not been.

A multitude of other solutions have been proposed. I will discuss just two ideas, which I regard as being on the right track.

Here is the first idea. Perhaps observing a white shoe or a black raven *may or may not* confirm "All ravens are black." It depends on other factors. Suppose we know, for some reason, that either (1) all ravens are black and ravens are extremely rare, or else (2) most ravens are black, a few are white, and ravens are common. Then a casual observation of a black raven will support (2), a hypothesis that says that not all ravens are black. If all ravens were black, we should not be seeing them at all. Observing a white shoe, similarly, may or may not confirm a given hypothesis, depending on what else we know. This reply was first suggested by I. J. Good (1967).

Good's move is very reasonable. We see here a connection to the issue of holism about testing, discussed in chapter 2. The relevance of an observation to a hypothesis is not a simple matter of the content of the two statements; it depends on other assumptions as well. This is so even in the simple case of a hypothesis like "All F's are G" and an observation like "Object A is both F and G." Good's point also reminds us how artificially simplified the standard logical empiricist examples are. No biologist would seriously wonder whether seeing thousands of black ravens makes it likely that all ravens are black. Our knowledge of genetics and bird coloration leads us to expect some variation, such as cases of albinism, even when we have seen thousands of black ravens and no other colors.

Here is a second suggestion about the ravens, which is consistent with Good's idea but goes further. Whether or not a black raven or a white shoe confirms "All ravens are black" might depend on the *order* in which you learn of the two properties of the object.

Suppose you hypothesize that all ravens are black, and someone comes up to you and says, "I have a raven behind my back; want to see what color it is?" You should say yes, because if the person pulls out a white raven, your theory is refuted. You need to find out what is behind his back. But suppose the person comes up and says, "I have a black object behind my back; want to see whether it's a raven?" Then it does not matter to you what is behind his back. You think that all ravens are black, but you don't have to think that all black things are ravens. In both cases, suppose the object behind his back is a black raven and he does show it to you. In the first situation, your observation of the raven seems relevant to your investigation of raven color, but in the other case it's irrelevant.

So perhaps the "All ravens are black" hypothesis is only confirmed by a black raven when this observation had the *potential to refute* the hypothesis, only when the observation was part of a genuine test.

Now we can see what to do with the white shoe. You believe that all ravens are black, and someone comes up and says, "I have a white object behind my back; want to see what it is?" You should say yes, because if he has a raven behind his back your hypothesis is refuted. He pulls out a shoe, however, so your hypothesis is OK. Then someone comes up and says, "I have a shoe behind my back; want to see what color it is?" In this case you need not care. It seems that in the first of these two cases, you have gained some support for the hypothesis that all ravens are black. In the second case you have not.

So perhaps some white-shoe observations *do* confirm "All ravens are black," and some black-raven observations *don't*. Perhaps there is only confirmation when the observations arise during a genuine test, a test that has the potential to disconfirm as well as confirm.

Hempel saw the possibility of a view like this. His responses to Good's argument and to the order-of-observation point were similar, in fact. He said he wanted to analyze a relation of confirmation that exists *just between a hypothesis and an observation itself,* regardless of extra information we might have, and regardless of the order in which observations were made. But perhaps Hempel was wrong here; there is *no such relation.* We cannot answer the question of whether an observation of a black raven confirms the generalization unless we know something about the way the observation was made and unless we make assumptions about other matters as well.

Hempel thought that some observations are just "automatically" relevant to hypotheses, regardless of what else is going on. That is true in the case of the deductive refutation of generalizations; no matter how we come to see a nonblack raven, that is bad news for the "All ravens are black" hypothesis. But what is true for deductive *dis*confirmation is not true for confirmation.

Clearly this discussion of order-of-observation does not entirely solve the ravens problem. *Why* does order matter, for example, and what if both properties are observed at once? I will return to this issue in chapter 14, using a more complex framework. Putting it briefly, we can only understand confirmation and evidence by taking into account the *procedures* involved in generating data. Or so I will argue.

I will make one more comment on the ravens problem. This one is a digression, but it does help illustrate what is going on. In psychology there is a famous experiment called the "selection task" (Wason and Johnson-Laird 1972). The experiment has been used to show that many people (including highly educated people) make bad logical errors in certain circumstances. The experimental subject is shown four cards with half of each card masked. The subject is asked to answer this question: "Which masks do you have to remove to know whether it is true that if there is a circle on the left of a card, there is a circle on the right as well?" See fig. 3.1 and try to answer the question yourself before reading the next paragraph.

Large majorities of people in many (though not all) versions of this experiment give the wrong answer. Many people tend to answer "only card A" or "card A and card C." The right answer is A and D. Compare this to the ravens problem; the problems have the same structure. I am sure Hempel would have given the right answer if he had been a subject in the four-card experiment, but the selection task might show something interesting about

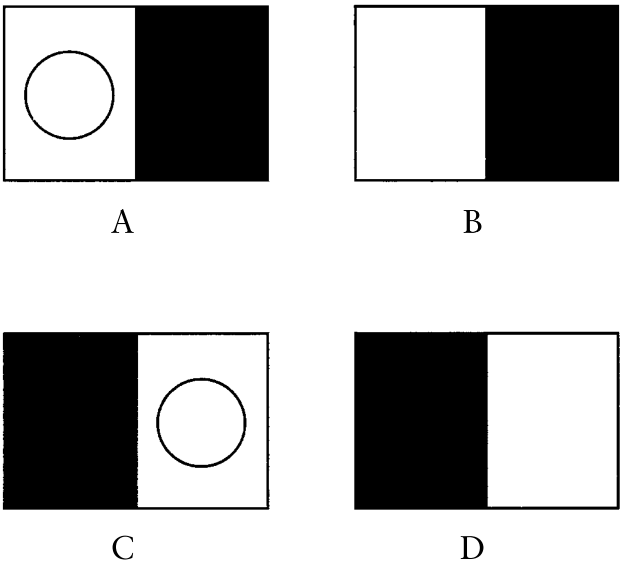

Fig. 3.1
The Wason selection task

why confirmation has been hard to analyze. For some reason it is difficult for people to see the importance of "card D" tests in cases like this, and it is easy for people to wrongly think that "card C" tests are important. If you are investigating the hypothesis that all ravens are black, card D is analogous to the situation when someone says he has a white object behind his back. Card C is analogous to the situation where he says he has a black object behind his back. Card D is a real test of the hypothesis, but card C is not. Unmasking Card C is evidentially useless, even though it may fit with what the hypothesis says. Not all observations of cases that fit a hypothesis are useful as tests.

3.4 Goodman's "New Riddle of Induction"

In this section I will describe an even more famous problem, revealed by Nelson Goodman (1955). This argument looks strange, and it is easy to misinterpret. But the issues it raises are very deep.

First we need to be clear about what Goodman was trying to do with his argument. His primary goal was to show that there cannot be a purely "formal" theory of confirmation. He does not think that confirmation is impossible, or that induction is a myth. He just thinks they work differently from the way many philosophers—especially logical empiricists—have thought.

What is a "formal" theory of confirmation? The easiest way to explain this is to look at deductive arguments. Recall the most famous deductively valid argument:

Argument 1

PREMISES	All men are mortal.
	Socrates is a man.
CONCLUSION	Socrates is mortal.

The premises, if they are true, guarantee the truth of the conclusion. But the fact that the argument is a good one does not have anything in particular to do with Socrates or manhood. Any argument that has the same form is just as good. That form is as follows:

All F's are G.
a is an F.

---

a is G.

Any argument with this form is deductively valid, no matter what we substitute for "F," "G," and "a." As long as the terms we substitute pick out definite properties or classes of objects, and as long as the terms retain the same meaning all the way through the argument, the argument will be valid.

So the deductive validity of arguments depends only on the form or pattern of the argument, not the content. This is one of the features of deductive logic that the logical empiricists wanted to build into their theory of induction and confirmation. Goodman aimed to show that this is impossible; there can never be a formal theory of induction and confirmation.

How did Goodman do it? Consider argument 2.

Argument 2

All the many emeralds observed, in diverse circumstances,
prior to 2010 A.D. have been green.

===

All emeralds are green.

This looks like a good inductive argument. (Like some of the logical empiricists, I use a double line between premises and conclusion to indicate that the argument is not supposed to be deductively valid.) The argument does not give us a guarantee; inductions never do. And if you would prefer to express the conclusion as "probably, all emeralds are green" that will not make any difference to the rest of the discussion.

(If you know something about minerals, you might object that emeralds are regarded as green by definition: emeralds are beryl crystals made green by trace amounts of chromium. Please just regard this as another unfortunate choice of example by the literature.)

Now consider argument 3:

Argument 3
All the many emeralds observed, in diverse circumstances,
prior to 2010 A.D. have been grue.

═══════════════════════════════════

All emeralds are grue.

Argument 3 uses a new word, "grue." We define "grue" as follows:

GRUE: An object is *grue* if and only if it was first observed before 2010 A.D. and is green, *or* if it was not first observed before 2010 A.D. and is blue.

The world contains lots of grue things; there is nothing strange about grue objects, even though there is something strange about the word. The grass outside my door as I write this is grue. The sky outside on July 1, 2020, will be grue, if it is a clear day. An individual object does *not* have to change color in order to be grue—this is a common misinterpretation. Anything green that has been observed before 2010 passes the test for being grue. So, all the emeralds we have seen so far have been grue.

Argument 3 does *not* look like a good inductive argument. Argument 3 leads us to believe that emeralds observed in the future will be blue, on the basis of previously observed emeralds being green. The argument also conflicts with argument 2, which looks like a good argument. But arguments 2 and 3 have *exactly the same form.* That form is as follows:

All the many E's observed, in diverse circumstances,
prior to 2010 A.D., have been G.

═══════════════════════════════════

All E's are G.

We could represent the form even more schematically than this, but that does not matter to the point. Goodman's point is that two inductive arguments can have the exact same form, but one argument can be good while the other is bad. So what makes an inductive argument a good or bad one cannot be just its form. Consequently, there can be no purely formal theory of induction and confirmation. Note that the word "grue" works perfectly well in *de*ductive arguments. You can use it in the form of argument 1, and it will cause no problems. But induction is different.

Suppose Goodman is right, and we abandon the idea of a formal theory of induction. This does not end the issue. We still need to work out *what* exactly is wrong with argument 3. This is the new riddle of induction.

The obvious thing to say is that there is something wrong with the word "grue" that makes it inappropriate for use in inductions. So a good theory of induction should include a *restriction* on the terms that occur in inductive arguments. "Green" is OK and "grue" is not.

This has been the most common response to the problem. But as Goodman says, it is very hard to spell out the details of such a restriction. Suppose we say that the problem with "grue" is that its definition includes a reference to a specific time. Goodman's reply is that whether or not a term is defined in this way depends on which language we take as our starting point. To see this, let us define a new term, "bleen."

BLEEN: An object is *bleen* if and only if it was first observed before 2010 A.D. and is blue, *or* if it was not first observed before 2010 A.D. and is green.

We can use the English words "green" and "blue" to define "grue" and "bleen," and if we do so we must build a reference to time into the definitions. But suppose we spoke a language that was like English except that "grue" and "bleen" were basic, familiar terms and "green" and "blue" were not. Then if we wanted to define "green" and "blue," we would need a reference to time.

GREEN: An object is *green* if and only if it was first observed before 2010 A.D. and is grue, *or* if it was not first observed before 2010 A.D. and is bleen.

(You can see how it will work for "blue.") So Goodman claimed that whether or not a term "contains a reference to time" or "is defined in terms of time" is a *language-relative* matter. Terms that look OK from the standpoint of one language will look odd from another. So if we want to rule out "grue" from inductions because of its reference to time, then whether an induction is good or bad will depend on *what language we treat as our starting point.* Goodman thought this conclusion was fine. A good induction, for Goodman, must use terms that have a history of normal use in our community. That was his own solution to his problem. Most other philosophers did not like this at all. It seemed to say that the value of inductive arguments depended on irrelevant facts about which language we happen to use.

Consequently, many philosophers have tried to focus not on the words "green" and "grue" but on the *properties* that these words pick out, or the *classes* or *kinds* of objects that are grouped by these words. We might argue that greenness is a natural and objective feature of the world, and grueness is not. Putting it another way, the green objects make up a "natural kind," a kind unified by real similarity, while the grue objects are an artificial or arbitrary collection. Then we might say: a good induction has to use terms that we have reason to believe pick out natural kinds. Taking this approach plunges us into hard problems in other parts of philosophy. What *is* a property? What *is* a "natural kind"? These are problems that have been controversial since the time of Plato.

Although Goodman's problem is abstract, it has interesting links to real problems in science. In fact, Goodman's problem encapsulates within it several distinct hard methodological issues in science; that is partly why the problem is so interesting. First, there is a connection between Goodman's problem and the "curve-fitting problem" in data analysis. Suppose you have a set of data points in the form of x and y values, and you want to discern a general relationship expressed by the points by fitting a function to them. The points in figure 3.2 fall almost exactly on a straight line, and that seems to give us a natural prediction for the y value we expect for $x = 4$. However, there is an infinite number of different mathematical functions that fit our three data points (as well or better) but which make different predictions for the case of $x = 4$. How do we know which function to use? Fitting a strange function to the points seems to be like preferring a grue induction over a green induction when inferring from the emeralds we have seen.

Scientists dealing with a curve-fitting problem like this may have extra information telling them what sort of function is likely here, or they may prefer a straight line on the basis of *simplicity.* That suggests a way in which we might deal with Goodman's original problem. Perhaps the green induction is to be preferred on the basis of its simplicity?

That might work, but there are problems. First, is it really so clear that the green induction is simpler? Goodman will argue that the simplicity of an inductive argument depends on which language we assume as our starting point, for the kinds of reasons given earlier in this section. For Goodman, what counts as a simple pattern depends on which language you speak or which categorization you assume. Also, though a preference for

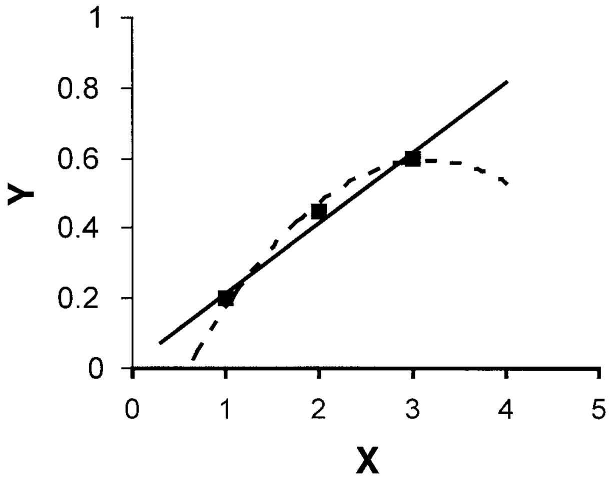

Fig. 3.2
The curve-fitting problem

simplicity is very common in science, such a preference is often hard to justify. Simpler theories are easier for us to work with, but that does not seem to give us reason to prefer them if we are seeking to learn what the world is really like. Why should the world be simple rather than complex?

Earlier I mentioned attempts to solve Goodman's problem using the idea of a "natural kind," a collection unified by real similarity as opposed to stipulation or convention. Though this term is philosophical, a lot of argument within science is concerned just this sort of problem—with getting the right *categories* for prediction and extrapolation. The problem is especially acute in sciences like economics and psychology that deal with complex networks of similarities and differences across the cases they try to generalize about. Do all economies with very high inflation fall into a natural kind that can be used to make general predictions? Are the mental disorders categorized in psychiatric reference books like the *DSM IV* really natural kinds, or have we applied standard labels like "schizophrenia" to groups of cases that have no real underlying similarity? The periodic table of elements in chemistry seems to pick out a set of real natural kinds, but is this something we can hope for in all sciences? If so, what does that tell us about inductive arguments in different fields?

That concludes our initial foray into the problems of induction and confirmation. These problems are simple, but they are very resistant to solution. For a good part of the twentieth century, it seemed that even the most innocent-looking principles about induction and confirmation led straight into trouble.

Later (especially in chapter 14) I will return to these problems. But in the next chapter we will look at a philosophy that gets a good part of its motivation from the frustrations discussed in this chapter.

Further Reading

Once again, Hempel's *Aspects of Scientific Explanation* (1965) is a key source, containing a long (and exhausting) chapter on confirmation. Skyrms, *Choice and Chance* (2000), is a classic introductory book on these issues, and it introduces probability theory as well. Even though it argues for a view that will not be discussed until chapter 14, Howson and Urbach's *Scientific Reasoning* (1993) is a useful introduction to various approaches to confirmation. It has the most helpful short summary of Carnap's ideas that I have read. Carnap's magnum opus on these issues is his *Logical Foundations of Probability* (1950). For a discussion of explanatory inference, see Lipton, *Inference to the Best Explanation* (1991).

For the use of order-of-observation to address the ravens problem, see Horwich, *Probability and Evidence* (1982), but you should probably read chapter 14 of this book first.

Goodman's most famous presentation of his "new riddle of induction" is in *Fact, Fiction & Forecast* (1955). The problem is in chapter 3 (along with other interesting ideas), and his solution is in chapter 4. His subsequent papers on the topic are collected in *Problems and Projects* (1972). Douglas Stalker has edited a collection on Goodman's riddle, called *Grue!* (1994). It includes a very detailed bibliography. The Quine and Jackson papers are particularly good.

For discussions of properties and kinds, and their relevance to induction, see Armstrong 1989, Lewis 1983, Dupre 1993, and Kornblith 1993. (These are fairly advanced discussions, except for Armstrong's, which is introductory.) There is a good discussion of simplicity in Sober 1988.

4

Popper: Conjecture and Refutation

4.1 Popper's Unique Place in the Philosophy of Science

Karl Popper is the only philosopher discussed in this book who is regarded as a hero by many scientists. Attitudes toward philosophy among scientists vary, but hardly ever does a philosopher succeed in *inspiring* scientists in the way Popper has. It is also rare for a philosopher's view of science to be used within a scientific debate to justify one position over another. This has happened with Popper too. Within biology, recent debates about the classification of organisms and about ecology have both seen Popper's ideas used in this way (Hull 1999). I once went to a lecture by a famous virologist who had won a Nobel Prize in medicine, to hear about his work. What I heard was mostly a lecture about Popper. In 1965, Karl Popper even became *Sir* Karl Popper, knighted by the queen of England.

Popper's appeal is not surprising. His view of science is centered around a couple of simple, clear, and striking ideas. His vision of the scientific enterprise is a noble and heroic one. Popper's theory of science has been criticized a great deal by philosophers over the years. I agree with many of these criticisms and don't see any way for Popper to escape their force. Despite the criticism, Popper's views continue to have an important place in philosophy and continue to appeal to many working scientists.

4.2 Popper's Theory of Science

Popper began his intellectual career in Vienna, between the two world wars. He was not part of the Vienna Circle, but he did have contact with the logical positivists. This contact included a lot of disagreement, as Popper developed his own distinctive position. Popper does count as an "empiricist" in the broad sense used in this book, but he spent a lot of time distinguishing his views from more familiar versions of empiricism. Like the logical positivists, Popper left Europe upon the rise of Nazism, and after spending the war years in New Zealand, he moved to the London School of Economics, where he remained for the rest of his career. There he built up a loyal group of allies, whom he often accused of disloyalty. His seminar series at the London School of Economics became famous for its grueling questioning and for the fact that speakers had a difficult time actually presenting much of their lectures, because of Popper's interruptions.

Popper once had a famous confrontation with Wittgenstein, on the latter's turf at Cambridge University. One version of the story, told by Popper himself, has Wittgenstein brandishing a fireplace poker during a discussion of ethical rules, leading Popper to give as an example of an ethical rule: "not to threaten visiting lecturers with pokers." Wittgenstein stormed out. Other versions of the story, including those told by Wittgenstein's allies, deny Popper's account (see Edmonds and Eidinow 2001 for this controversy).

The logical positivists developed their theory of science as part of a general theory of language, meaning, and knowledge. Popper was not much interested in these broader topics, at least initially; his primary aim was to understand science. As his first order of business, he wanted to understand the difference between scientific theories and nonscientific theories. In particular, he wanted to distinguish science from "pseudo-science." Unlike the logical positivists, he did not regard pseudo-scientific ideas as meaningless; they just weren't science. For Popper, an inspiring example of genuine science was the work of Einstein. Examples of pseudo-science were Freudian psychology and Marxist views about society and history.

Popper called the problem of distinguishing science from non-science the "problem of demarcation." All of Popper's philosophy starts from his proposed solution to this problem. "Falsificationism" was the name Popper gave to his solution. Falsificationism claims that *a hypothesis is scientific if and only if it has the potential to be refuted by some possible observation*. To be scientific, a hypothesis has to take a risk, has to "stick its neck out." If a theory takes no risks at all, because it is compatible with every possible observation, then it is not scientific. As I said above, Popper held that Marx's and Freud's theories were not scientific in this sense. No matter what happens, Popper thought, a Marxist or a Freudian can fit it somehow into his theory. So these theories are never exposed to any risks.

So far I have described Popper's use of falsifiability to distinguish scientific from nonscientific theories. Popper also made use of the idea of falsification in a more far-reaching way. He claimed that *all* testing in science has the form of attempting to refute theories by means of observation. And crucially, for Popper it is never possible to confirm or establish a theory by showing its agreement with observations. *Confirmation is a myth*. The only thing an observational test can do is to show that a theory is false. So the truth of a scientific theory can never be supported by observational evidence, not even a little bit, and not even if the theory makes a huge number of predictions that all come out as expected.

As you might think, Popper was a severe critic of the logical empiricists' attempts to develop a theory of confirmation or "inductive logic." The problems they encountered, some of which I discussed in chapter 3, were music to his ears. Popper, like Hume, was an inductive skeptic, and Popper was skeptical about all forms of confirmation and support other than deductive logic itself.

Skepticism about induction and confirmation is a much more controversial position than Popper's use of falsification to solve the demarcation problem. Most philosophers of science have thought that if induction and confirmation are just myths, that is very bad news for science. Popper tried to argue that there is no reason to worry; induction is a myth, but science does not need it anyway. So inductive skepticism, for Popper, is no threat to the rationality of science. In the opinion of most philosophers, Popper's attempt to defend this radical claim was not successful, and some of his discussions of this topic are rather misleading to readers. As a result, some of the scientists who regard Popper as a hero do not realize that Popper believed it is never possible to confirm a theory, not even slightly, and no matter how many observations the theory predicts successfully.

Popper placed great emphasis on the idea that we can never be *completely sure* that a theory is true. After all, Newton's physics was viewed as the best-supported theory ever, but early in the twentieth century it was shown to be false in several respects. However, almost all philosophers of science accept that we can never be 100 percent certain about factual matters, especially those discussed in science. This position, that we can never be completely certain about factual issues, is often known as *fallibilism* (a term due to C. S. Peirce). Most philosophers of science accept fallibilism. The harder question is whether or not we can be reasonable in *increasing* our confidence in the truth of a theory when it passes observational tests. Popper said no. The logical empiricists and most other philosophers of science say yes.

So Popper had a fairly simple view of how testing in science proceeds. We take a theory that someone has proposed, and we deduce an observational prediction from it. We then check to see if the prediction comes out as the theory says it will. If the prediction fails, then we have refuted, or falsified, the theory. If the prediction does come out as predicted, then all we should say is that *we have not yet falsified the theory.* For Popper, we cannot conclude that the theory is true, or that it is probably true, or even that it is more likely to be true than it was before the test. The theory *might* be true, but we can't say more than that.

We then try to falsify the theory in some other way, with a new prediction. We keep doing this until we have succeeded in falsifying it. What if years pass and we seem to never be able to falsify a theory, despite repeated tests? We can say that the theory has now survived repeated attempts to falsify it, but that's all. We never increase our confidence in the truth of the theory; and ideally, we should never stop trying to falsify it. That's not to say we should spend all our time testing theories that have passed tests over and over again. We do not have the time and resources to test everything that could be tested. But that is just a practical constraint. According to Popper, we should always retain a *tentative* attitude toward our theories, no matter how successful they have been in the past.

In defending this view, Popper placed great emphasis on the difference between confirming and disconfirming statements of scientific law. If someone proposes a law of the form "All F's are G," all it takes is one observation of an F that is not a G to falsify the hypothesis. This is a matter of deductive logic. But it is never possible to assemble enough observations to conclusively demonstrate the truth of such a hypothesis. You might wonder about situations where there is only a small number of F's and we could hope to check them all. Popper and the logical empiricists regarded these as unimportant situations that do not often arise in science. Their aim was to describe testing in situations where there is a huge or infinite number of cases covered by a hypothesized law or generalization. So Popper stressed that universal statements are hard or impossible to verify but easy, in principle, to falsify. The logical empiricist might reply that statements of the form "Some F's are G" have the opposite feature; they are easy to verify but hard or impossible to falsify. But Popper claimed (and the logical empiricists tended to agree) that real scientific theories rarely take this form, even though some statements in science do.

Despite insisting that we can never support or confirm scientific theories, Popper believed that science is a search for true descriptions of the world. How can one search for truth if confirmation is impossible?

This is an unusual kind of search. We might compare it to a certain kind of search for the Holy Grail, conducted by an imaginary medieval knight. Suppose there are lots of grails around, but only one of them is holy. In fact, the number of nonholy grails is infinite or enormous, and you will never encounter them all in a lifetime. All the grails glow, but only the Holy Grail glows forever. The others eventually stop glowing, but there is no telling when any particular nonholy grail will stop glowing. All you can do is pick up one grail and carry it around and see if it keeps on glowing. You are only able to carry one at a time. If the one you are carrying is the Holy Grail, it will never stop glowing. But you would never *know* if you currently had the Holy Grail, because the grail you are carrying might stop glowing at any moment. All you can do is reject grails that are clearly not holy (since they stop glowing at some point) and keep picking up a new one. You will eventually die (with no afterlife, in this scenario) without knowing whether you succeeded.

This is similar to Popper's picture of science's search for truth. All we can do is try out one theory after another. A theory that we have failed to falsify up till now *might,* in fact, be true. But if so, we will never know this or even have reason to increase our confidence.

4.3 Popper on Scientific Change

So far I have described Popper's views about the demarcation of science from non-science and the nature of scientific testing. Popper also used the idea of falsification to propose a theory of scientific *change.*

Popper's theory has an appealing simplicity. Science changes via a two-step cycle that repeats endlessly. Stage 1 in the cycle is *conjecture*—a scientist will offer a hypothesis that might describe and explain some part of the world. A good conjecture is a *bold* one, one that takes a lot of risks by making novel predictions. Stage 2 in the cycle is *attempted refutation*—the hypothesis is subjected to critical testing, in an attempt to show that it is false. Once the hypothesis is refuted, we go back to stage 1 again—a new conjecture is offered. That is followed by stage 2, and so on.

As the process moves along, it is natural for a scientist to propose conjectures that have some relation to previous ones. A theoretical idea can be refined and modified via many rounds of conjecture and refutation. That is fine, for Popper, though it is not essential. One thing that a scientist should *not* do, however, is to react to the falsification of one conjecture by cooking up a new conjecture that is designed to just avoid the problems revealed by earlier testing, and which goes no further. We should not make *ad hoc* moves that merely patch the problems found in earlier conjectures. Instead, a scientist should constantly strive to *increase* the breadth of application of a theory and increase the precision of its predictions. That means constantly trying to increase the "boldness" of conjectures.

What sort of theory is this? Popper intended it as a description of the general pattern that we *actually* see in science, and as a description of *good* scientific behavior as well. He accepted that not all scientists succeed in sticking to this pattern of behavior all the time. Sometimes people become too wedded to their hypotheses; they refuse to give them up when testing tells them to. But Popper thought that a lot of actual scientific behavior does follow this pattern and that we see it especially in great scientists such as Einstein. For Popper, a good or great scientist is someone who combines two features, one corresponding to each stage of the cycle. The first feature is an ability to come up with imaginative, creative, and risky ideas. The second is a hard-headed willingness to subject these imaginative ideas to rigorous critical testing. A good scientist has a creative, almost artistic, streak and a tough-minded, no-nonsense streak. Imagine a hard-headed cowboy out on the range, with a Stradivarius violin in his saddlebags. (Perhaps at this point you can see some of the reasons for Popper's popularity among scientists.)

Popper's view here can apparently be applied in the same way to *individuals* and to *groups* of scientists. An isolated individual can behave scientifically by engaging in the process of conjecture and refutation. And a collection of scientists can each, at an individual level, follow Popper's two-step procedure. But another possibility is a division of labor; one individual (or team) comes up with a conjecture, and another does the attempted refutation. Popper's basic description of the two-step conjecture-and-refutation pattern of science seems compatible with all these possibilities. But the case where individual A does the conjecture and individual B does the refutation will be suspicious to Popper. If individual A is a true scientist, he should take a critical attitude toward his own ideas. If individual A is completely fixated on his conjecture, and individual B is fixated on showing that A is wrong in order to advance his *own* conjecture, this is not good scientific behavior according to Popper.

This raises an interesting question. Empiricist philosophies stress the virtues of open-mindedness, and Popper's view is no exception. But perhaps an open-minded *community* can be built out of a collection of rather closed-minded *individuals*. If actual scientists are wedded to their own conjectures, but each is wedded to a *different* conjecture and would like to prove the others wrong, shouldn't the overall process of conjecture and refutation work? What is wrong with the situation where B's role is to critically test A's ideas? So long as the testing occurs, what does it matter whether A or B does it? One problem is that if everyone is so closed-minded, the results of the test might have no impact on what people believe. Perhaps the young and tender minds of incoming graduate students could be the community's source of flexibility; unsuccessful theories will attract no new recruits and will die with their originators. This would be a rather slow way for science to change (but many would argue that we do see cases like this).

In later chapters of this book, we will look at theories that focus on social structure in science, and at various kinds of division of labor between individual scientists. Although Popper did stress community standards in science, he did seem to have a picture in which the good scientist should, *as an individual,* have the willingness to perform both the imaginative and the critical roles. A good scientist should retain a tentative attitude toward all theories, including his own.

I will make one more point before moving on to criticisms of Popper. The two-step process of conjecture and refutation that Popper describes has a striking resemblance to another two-step process: Darwin's explanation of biological evolution in terms of *variation and natural selection.* In science according to Popper, scientists toss out conjectures that are subjected to critical testing. In evolution, according to both Darwin himself and more recent versions of evolutionary theory, populations evolve via a process in which variations appear in organisms in a random or "undirected" way, and these novel characteristics are "tested" through their effects on the organism in its interactions with the environment. Variations that help organisms to survive and reproduce, and which are of the kind that gets passed on in reproduction, tend to be preserved and become more common in the population over time.

Ironically, at one time Popper thought that Darwinism is not a scientific theory, but he later retracted that claim. In any case, both Popper and others have explored the analogy between Popperian science and Darwinian evolution in detail. The analogy should not be taken *too* seriously; evolution is not a process in which populations really "search" for anything, in the way that scientists search for good theories, and there are other crucial differences too. But the similarity is certainly interesting. Analogies between science and evolution will come again in later chapters (6 and 11).

4.4 Objections to Popper on Falsification

Let us now turn to a critical assessment of Popper's ideas. We should start with his solution to the demarcation problem. Is falsifiability a good way to distinguish scientific ideas from nonscientific ones?

Let me first say that I think this question probably has no answer in the form in which Popper expressed it. We should not expect to be able to go through a list of statements or theories and label them "scientific" or "not scientific." However, I suggest that something fairly similar to Popper's question about demarcation does make sense: can we describe a distinctive scientific *strategy* of investigating the world, a scientific way of *handling* ideas?

Some of Popper's ideas are useful in trying to answer this question. In particular, Popper's claim that scientific theories should take *risks* is a good one; this will be followed up in the last section of this chapter. But Popper had an overly simple picture of *how* this risk-taking works.

For Popper, theories have the form of generalizations, and they take risks by prohibiting certain kinds of particular events from being observed. If we believe that all pieces of iron, of whatever size and shape, expand when heated, then our theory forbids the observation of something that we know to be a piece of iron contracting when heated. A problem may have occurred to you: how sure can we be that, if we see a piece of "iron" contracting when heated, that it is really iron? We might also have doubts about our measurements of the contraction and the temperature change. Maybe the generalization about iron expanding when heated is true, but our assumptions about the testing situation and our ability to know that a sample is made of iron are false.

This problem is a reappearance of an issue discussed in chapter 2: holism about testing. Whenever we try to test a theory by comparing it with observations, we must make a large number of additional assumptions in order to bring the theory and the observations into "contact" with each other. If we want to test whether iron always expands when heated, we need to make assumptions about our ability to find or make reasonably pure samples of iron. If we want to test whether the amounts of the bases C and G are equal and the amounts of A and T are equal in all samples of DNA (Chargaff's rules), we need to make a lot of assumptions about our chemical techniques. If we observe an unexpected result (iron contracting on heating, twice as much C as G in a sample of DNA), it is always possible to blame one of these extra assumptions rather than the theory we are trying to test. In extreme cases, we might even claim that the apparent observation was completely misunderstood or wrongly described by the observers. Indeed, this is not so uncommon in our attempts to work out what to make of reports of miracles and UFO abductions. So how can we really use observations to falsify theories in the way Popper wants?

This is a problem not just for Popper's solution to the demarcation problem, but for his whole theory of science as well.

Popper was well aware of this problem, and he struggled with it. He regarded the extra assumptions needed to connect theories with testing situations as scientific claims that might well be false—these are conjectures too. We can try to test these conjectures separately. But Popper conceded that logic itself can never *force* a scientist to give up a particular theory, in the face of surprising observations. Logically, it is always possible to blame other assumptions involved in the test. Popper thought that a good scientist would not try to do this; a good scientist is someone who wants to expose the theory itself to tests and will not try to deflect the blame.

Does this answer the holist objection? What Popper has done is to move from describing a characteristic of scientific *theories* to describing a characteristic of scientific *behavior.* In some ways this is a retraction of his initial aim, which was to describe something about scientific theories themselves that makes them special. That is a problem. Then again, this shift to describing scientific modes of thought and behavior, rather than theories, may well be a step forward. This will be discussed in more detail in section 4.6.

Popper also accepted that we cannot be completely certain about the observation reports that we use to falsify theories. We have to regard the acceptance of an observation report as a "decision," one that is freely made. Once we have made the decision, we can use the observation report to falsify any theory that conflicts with it. But for Popper, any falsification process is based, in the end, on a decision that could be challenged. Someone might come along later and try to show, via more testing, that the observation report was not a good one—that person might investigate whether the conditions of observation were misleading. That testing has the same conjecture-and-refutation form described earlier. So this investigation into the controversial observation ultimately depends on "decisions" too.

Is this bad news for Popper? Popper insisted that making these decisions about single observations is very different from making free decisions directly about the theories themselves. But what *sort* of difference is this? If observation reports rest on nothing more than "decisions," and these determine our choice of theories, how is that better than directly choosing the theories themselves, without worrying about observation? Or why couldn't we just "decide" to hang onto a theory and reject the observation reports that conflict with it? I am not saying that we *should* do these things, just that Popper has not given us a good reason not to do them. I believe that we should not do these things because we have good reason to believe that observation is a generally reliable way of forming beliefs. As I will argue in chapter 10, we need to make use of a scientific theory of perception at this point in the story. But that argument will have to come later. Popper himself does not try to answer these questions by giving an argument about the reliability of perception.

This point about the role of decisions affects Popper's ideas about demarcation as well as his ideas about testing. Any system of hypotheses can be held onto despite apparent falsification, if people are willing to make certain decisions. Does that mean that Popper's theory fails to differentiate between science and pseudo-science after all? The answer is "yes and no." The yes comes from the fact that scientific theories can be handled in a way that makes them immune to falsification, and nonscientific theories can be rejected if people decide to accept claims about particular matters that are incompatible with the theory. But there is a "no" part in the answer as well. A scientific theory is falsifiable via a certain *kind* of decision—a decision about an observation report. A pseudo-scientific theory, Popper says, does not clash with any possible observations. So if a pseudo-scientific theory is to be rejected, some different kind of decision must be made. We can accept, with Popper, that this is a significant difference. But Popper has not told us why this way of doing things, the scientific way, is more rational than some other way.

I have been fairly tough on Popper's views about falsification in this section, and there is another problem to discuss as well. The problem is bad for Popper, but I should emphasize that it is bad for many others as well.

What can Popper say about theories that do not claim that some observation O is forbidden, but only that it is very *unlikely*? If I believe that a certain coin is "fair," I can deduce from this hypothesis various claims about the probabilities of long "all heads" or "all tails" sequences of tosses. Suppose I observe 100 tosses turning up heads 100 times. This is very unlikely according to my hypothesis about the coin, but it is not impossible. *Any* finite stretch of heads tosses is possible with a fair coin, although longer and longer runs of heads are treated by the theory as more and more unlikely. But if a hypothesis does not *forbid* any particular observations, then, according to Popper, it is taking no risks. That seems to entail that theories that ascribe low probabilities to specific observations, but do not rule them out altogether, are unfalsifiable and hence unscientific for Popper.

Popper's response was to accept that, logically speaking, all hypotheses of this kind *are* unscientific. But this seems to make a mockery of the important role of probability in science. So Popper said that a scientist can decide that if a theory claims that a particular observation is extremely improbable, the theory *in practice* rules out that observation. So if the observation is made, the theory is, in practice, falsified. According to Popper, it is up to scientists to work out, for their own fields, what sort of probability is so low that events of that kind are treated as prohibited. So probabilistic theories can only be construed as falsifiable in a special "in practice" sense. And we have here another role for "decisions" in Popper's philosophy of science, as opposed to the constraints of logic.

Popper is right that scientists reject theories when observations occur which the theory says are highly improbable (although it is a complicated matter which *kinds* of improbability have this importance). And Popper is right that scientists spend a good deal of time working out "how improbable is *too* improbable." Complex statistical methods are used to help scientists with these decisions. But in making this move, Popper has badly damaged his original picture of science. This was a picture in which observations, once accepted, have the power to decisively refute theoretical hypotheses. That is a matter of deductive logic, as Popper endlessly stressed. Now Popper is saying that falsification can occur without its being backed up by a deductive logical relation between observation and theory.

4.5 Objections to Popper on Confirmation

As described earlier, Popper believed that theories are never confirmed by observations, and he thought inductive arguments are never justified. Popper thought that a theory of the rational choice of theories could be given entirely in terms of falsification, so he thought that rejecting induction and confirmation was no problem.

In the previous section I discussed problems with Popper's views about falsification. But let us leave those problems aside now, and assume in this section that we can use Popperian falsification as a method for decisively rejecting theories. If we make this assumption, is Popper's attempt to describe rational theory choice successful? No, it is not.

Here is simple problem that Popper has a very difficult time with. Suppose we are trying to build a bridge, and we need to use physical theories to tell us which designs are stable and will support the weight that the bridge must carry. This is a situation where we must apply our scientific theories to a practical task. As a matter of fact, engineers and scientists in this situation will undoubtedly tend to use physical theories that have survived empirical testing; they will use "tried and true" methods as far as possible. The empiricist approach to the philosophy of science holds that such a policy is rational. The problem for an empiricist philosophy is to explain in more detail *why* this policy is the right one. That task is hard, as I hope became clear in chapter 3. But let us focus on Popper, who wants to avoid the need for a theory of confirmation. How does Popper's philosophy treat the bridge-building situation?

Popper can say why we should prefer to use a theory that has not been falsified over a theory that has been falsified. Theories that have been falsified have been shown to be false (here again I ignore the problems discussed in the previous section). But suppose we have to choose between (1) a theory that has been tested many times and has passed every test, and (2) a brand new theory that has just been conjectured and has never been tested. Neither theory has been falsified. We would ordinarily think that the rational thing to do is to choose the theory that has survived testing. But what can Popper say about this choice? Why exactly would it be irrational, for Popper, to build the bridge using a brand new theory that has never been tested?

Popper recognized and struggled with this problem too. Perhaps this has been the most common objection to Popper from other empiricist philosophers (e.g., Salmon 1981). Popper is not able to give a very good reply.

Popper refuses to say that when a theory passes tests, we have more reason to believe that the theory is true. Both the untested theory and the well-tested theory are just conjectures. But Popper did devise a special concept to use in this situation. Popper said that a theory that has survived many attempts to falsify it is "corroborated." And when we face choices like the bridge-building one, it is rational to choose corroborated theories over theories that are not corroborated.

What is "corroboration"? Popper gave a technical definition and held that we can measure the amount of corroboration that a theory has at a particular time. The technicalities do not matter, though. We need to ask, What *sort* of property is corroboration? Has Popper just given a new name to confirmation? If so, he can answer the question about building the bridge, but he has given up one of his main differences from the logical empiricists and everyone else. If corroboration is totally different from confirmation—so different that we cannot regard corroboration as any guide to *truth*—then why should we choose a corroborated theory when we build the bridge? This issue has been much discussed (see Newton-Smith 1981). Popper's concept of corroboration can be interpreted in a way that makes it different from confirmation, but Popper can give no good answer to why we should choose corroborated theories over new ones when building bridges.

To understand corroboration, think of the difference between an academic transcript and a letter of recommendation. This distinction should be vivid to students! An academic transcript says what you have *done*. It measures your past performance, but it does not contain explicit predictions about what you will do in the future. A letter of recommendation usually says something about what you have done, and it also makes claims about how you are likely to do in the future. Confirmation, as understood by the logical empiricists, is something like a letter of recommendation for a scientific theory. Corroboration, for Popper, is only like an academic transcript. And Popper thought that no good reasons could be given for believing that past performance is a reliable guide to the future. So corroboration is entirely "backward-looking." Consequently, no reason can be given for building a bridge with a corroborated theory rather than a noncorroborated but unfalsified one.

I think the best thing for Popper to say about the bridge-building situation is to stick to his inductive skepticism. He should argue that we really *don't* know what will happen if we build another bridge with a design that has worked in the past. Maybe it will stay up and maybe it won't. There might also be practical reasons for choosing that design if we are very familiar with it. But if someone comes along with a brand new untested design, we won't know whether it's a bad design until we try it.

Popper liked to say here that there is no alternative policy that is *more* rational than using the familiar and well-tested design, and we do have to make *some* decision. So we can go ahead and use the established design. But as Wesley Salmon (1981) replied, this does not help at all. If confirmation does not exist, then it seems there is also no policy that is *more* rational than choosing the *un*tested design. All we have here is a kind of "tie" between the options.

For most people, this is an unsatisfactory place for a philosophy of science to end up. Inductive skepticism of this kind is hard to take seriously outside of abstract, academic discussion. However, the efforts of the last two hundred years have shown how extremely hard it is to produce a good theory of induction and confirmation. One of the valuable roles of Popper's philosophy is to show what sort of theory of science might be possible if we give up on induction and confirmation.

In the first chapter of this book, I said that few philosophers still try to give descriptions of a definite "scientific method," where this is construed as something like a recipe for science. Popper is something of an exception here, since he does come close to giving a kind of recipe (although Popper insists there is no recipe for coming up with interesting conjectures). His view has an interesting relationship to descriptions of scientific method given in science textbooks.

In many textbooks, one finds something called the "hypothetico-deductive method." Back in chapter 3, I discussed a view about confirmation that is often called "hypothetico-deductivism." Now we are dealing with a method rather than a theory of confirmation. Science textbooks are more cautious about laying out recipes for science than they used to be, but descriptions of the hypothetico-deductive method are still fairly common. Formulations of the method vary, but some are basically a combination of Popper's view of testing and a less skeptical view about confirmation. In these versions, the hypothetico-deductive method is a process in which scientists come up with conjectures and then deduce observational predictions from those conjectures. If the predictions come out as the theory says, then the theory is supported. If the predictions do not come out as the theory says, the theory is not supported and should be rejected.

This process has the basic pattern that Popper described, but the idea that theories can be "supported" by observations is *not* a Popperian idea.

The term "support" is vague, but I think discussions of the hypothetico-deductive method generally assume that if a theory makes a lot of successful predictions, we have more reason to believe that the theory is true than we had before the successful predictions were made. We will never be completely sure, of course. But the more tests a theory passes, the more confidence we can have in its truth. The idea that we can gradually increase our confidence that a theory is true is an idea that Popper rejected. As I said at the start of this chapter, some of Popper's scientific admirers do not realize that Popper's view has this feature, because some of Popper's discussions were misleading.

Other formulations of the hypothetico-deductive method include a first stage in which observations are collected and a conjecture is generated *from* these observations. Popper disagreed with this picture of scientific procedure because he argued that fact-gathering can only take place in a way guided by a conjecture. But this is a fairly minor point.

Another term that some textbooks use in discussing scientific method (though not so much any more) is "strong inference." This term was introduced by a chemist named John Platt (1964). Strong inference is roughly a Popperian kind of testing plus another further assumption, which Popper rejected. This assumption is that we can write down *all* the possible theories that might be true in some area, and test them one by one. We find the true theory by eliminating the alternatives—it's a kind of "Sherlock Holmes" method. For Popper, this is impossible. In any real case, there will be an infinite number of competing theories. So even if we eliminate ten or one hundred possibilities, there is still the same infinite number remaining. According to Popper, all we can do is to choose one theory, test it, then choose another, and so on. We can never have confidence that we have eliminated all, or most, of the alternatives. (More recent attempts to make use of this "Sherlock Holmes" method will be discussed in chapter 14.)

I have not discussed objections to Popper's theory of scientific change (section 4.3) yet, but I will do so in the next few chapters.

What is Popper's single most important and enduring contribution to philosophy of science? I'd say it is his use of the idea of "riskiness" to describe the kind of contact that scientific theories have with observation. Popper was right to concentrate on the ideas of exposure and risk in his description of science. Science tries to formulate and handle ideas in such a way that they are exposed to falsification and modification via observation. Popper's formulation is valuable because it captures the idea that theories can *appear* to have lots of contact with observation when in fact they only have a kind of "pseudo-contact" with observation because they are exposed to no risks. This is an advance in the development of empiricist views of science. Popper's analysis of *how* this exposure works does not work too well, but the basic idea is good.

4.6 Further Comments on the Demarcation Problem

Popper is onto something when he says that scientific theories should take risks. In this section I will try to develop this idea a bit differently.

Popper was interested in distinguishing scientific theories from unscientific ones, and he wanted to use the idea of risk-taking to make the distinction. But this idea of risk-taking is better used as a way of distinguishing scientific from unscientific ways of *handling* ideas. And we should not expect a sharp distinction between the two.

The scientific way of handling an idea is to try to connect it with other ideas, to embed it in a larger conceptual structure, in a way that *exposes it to observation*. This "exposure" is not a matter of simple falsification; there are many ways in which exposure to observation can be used to modify and assess an idea. But if a hypothesis is handled in a way that keeps it apart from all the risks associated with observation, that is an unscientific handling of the idea.

So it is a mistake to try to work out whether theories like Marxism or Freudianism are themselves "scientific" or not, as Popper did. A big idea like Marxism or Freudianism will have scientific and unscientific *versions,* because the main principles of the theory can be handled scientifically or unscientifically. Scientific versions of Marxism and Freudianism are produced when the main principles are connected with other ideas in a way that exposes these principles to testing. To scientifically handle the basic principles of Marxism is to try to work out what *difference* it would make to things we can observe if the Marxist principles were true. To do this it is not necessary that we write down some single observation that, if we observe it, will lead us to definitively reject the main principles of the theory. It will remain possible that an auxiliary assumption is at fault, and there is no simple recipe for adjudicating such decisions.

To continue with Popper's examples, Marxism holds that the driving force of human history is struggle between economic classes, guided by ongoing changes in economic organization. This struggle results in a predictable sequence of political changes, leading eventually to socialism. Freudianism holds that the normal development of a child includes a series of interactions and conflicts between unconscious aspects of the child's mind, where these interactions have a lot to do with resolving sexual feelings toward his or her parents. Adventurous ideas like these can be handled scientifically or unscientifically. Over the twentieth century, the Marxist view of history has been handled scientifically enough for it to have been disconfirmed. Too much has happened that seems to have little to do with class struggle; the ever-increasing political role of religious and cultural solidarity is an example (Huntington 1996). And capitalist societies have adapted to problems—especially economic tensions—in ways that Marxist views about politics and economics do not predict. Of course, it remains *possible* to hang onto the main principles of Marxism despite this, but fewer and fewer people handle the theory in that way anymore. Many still think that Marxism contains useful insights about economic matters, but the fundamental claims of the theory have not stood up well.

Freudianism is another matter; the ideas are still popular in some circles, but not because of success under empirical testing. Instead, the theory seems to hang around because of its striking and intriguing character, and because of a subculture in fields such as psychotherapy and literary theory which guards the main ideas and preserves them despite their empirical problems. The theory is handled very unscientifically by those groups. Freud's theory is not taken seriously by most scientifically oriented psychology departments in research universities, but it is taking a while for this fact to filter out to other disciplines.

Evolution is another big idea that can be handled scientifically or unscientifically. People (including Popper) have wondered from time to time whether evolutionary theory, or some specific version of it such as Darwinism, is testable. So they have asked, What observations would lead scientists to give up current versions of evolutionary theory? A one-line reply that biologists sometimes give to this question is "a Precambrian rabbit." An evolutionary biology textbook by Douglas Futuyma expresses the same point more soberly: finding "incontrovertibly mammalian fossils in incontrovertibly pre-cambrian rocks" would "refute or cast serious doubt on evolution" (1998, 760). The one-liner is a start, but the real situation is more complicated. So let us look at the case.

The Precambrian era ended around 540 million years ago. Suppose we found a well-preserved rabbit fossil in rocks 600 million years old. All our other evidence suggests that the only animals around then were sponges and a few other invertebrates and that mammals did not appear until over 300 million years later. Of course, a good deal of suspicion would be directed toward the finding itself. How sure are we that the rocks are that old? Might the rabbit fossil have been planted as a hoax? Remember the apparent fossil link between humans and apes that turned out to be a hoax, the Piltdown man of 1908 (see Feder 1996). Here we encounter another aspect of the problem of holism about testing—the challenging of observation reports, especially observation reports that are expressed in a way that presupposes other pieces of theoretical knowledge. This will be discussed in chapter 10. But let us suppose that all agree the fossil is clearly a Precambrian rabbit.

This finding would not be an instant falsification of all of evolutionary theory, because evolutionary theory is now a diverse package of ideas, including abstract theoretical models as well as claims about the actual history of life on earth. The theoretical models are intended to describe what various evolutionary mechanisms can do *in principle*. Claims of that kind are usually tested via mathematical analysis and computer simulation. Small-scale evolution can also be observed directly in the lab, especially in bacteria and fruit flies, and the Precambrian rabbit would not affect those results.

But a Precambrian rabbit fossil would show that *somewhere* in the package of central claims found in evolutionary biology textbooks, there are some very serious errors. These would at least include errors about the overall history of life, about the kinds of processes through which a rabbit-like organism could evolve, and about the "family tree" of species on earth. The challenge would be to work out where the errors lie, and that would require separating out and independently reassessing each of the ideas that make up the package. This reassessment could, in principle, result in the discarding of very basic evolutionary beliefs—like the idea that humans evolved from nonhumans.

Over the past twenty years or so, evolutionary theory has in fact been exposed to a huge and sustained empirical test, because of advances in molecular biology. Since the time of Darwin, biologists have been trying to work out the total family tree linking all species on earth, by comparing their similarities and differences and taking into account factors such as geographical distribution. The family tree that was arrived at prior to the rise of molecular biology can be seen summarized in various picturesque old charts and posters.

Then more recently, molecular biology made it possible to compare the DNA sequences of many species. Similarity in DNA is a good indicator of the closeness of evolutionary relationship. Claims about the evolutionary relationships between different species can be tested reasonably directly by discovering how similar their DNA is and calculating how many years of independent evolution the species have had since they last shared a "common ancestor." As this work began, it was reasonable to wonder whether the wealth of new information about DNA would be compatible or incompatible with the family tree that had been worked out previously. Suppose the DNA differences between humans and chimps had suggested that the human lineage split off from the lineage that led to chimps many hundreds of millions of years ago and that humans are very closely genetically related to squid. This would have been a disaster for evolutionary theory, one of almost the same magnitude as the Precambrian rabbit.

As it happened, the DNA data suggest that humans and chimps diverged about 4.6–5 million years ago and that chimps or pigmy chimps (bonobos) are our nearest living relatives. Prior to the DNA data, it was unclear whether humans were more closely related to chimps or to gorillas, and the date for the chimp-human divergence was much less clear. That is how the grand test of our old pre-molecular family tree has tended to go. There have been no huge surprises but lots of new facts and a lot of adjustments to the previous picture.

Further Reading

Popper's most famous work is his book *The Logic of Scientific Discovery,* published in German in 1935 and in English in 1959. The book is mostly very readable. Chapters 1–5 and 10 are the key ones. For the issues in section 4.4 above, see chapter 5 of Popper; for section 4.5, see chapter 10. A quicker and very useful introduction to Popper's ideas is the paper "Science: Conjectures and Refutations" in his collection *Conjectures and Refutations* (1963).

Newton-Smith, *The Rationality of Science* (1981), contains a clear and detailed assessment of Popper's ideas. It includes a simplified presentation of some of the technical issues surrounding corroboration that I omitted here. Salmon 1981 is an exceptionally good critical discussion of Popper's views on induction and prediction. See also Putnam 1974. Schilpp (1974) collects many critical essays on Popper, with Sir Karl's replies.

Popper's influence on biologists and his (often peculiar) ideas about evolutionary theory are discussed in Hull 1999. Horgan's book *The End of Science* (1996) contains a very entertaining interview with Popper.