Chapter 2
Scientific inference

Scientists often tell us things about the world that we would not otherwise have believed. For example, biologists tell us that we are closely related to chimpanzees, geologists tell us that Africa and South America used to be joined together, and cosmologists tell us that the universe is expanding. But how did scientists reach these unlikely sounding conclusions? After all, no one has ever seen one species evolve from other, or a single continent split into two, or the universe getting bigger. The answer, of course, is that scientists arrived at these beliefs by a process of *reasoning* or *inference*. But it would be nice to know more about this process. What exactly is the nature of scientific inference?

Deduction and induction

Logicians make an important distinction between *deductive* and *inductive* inference, or deduction and induction for short. An example of a deductive inference is the following:

> All Frenchmen like red wine
> Pierre is a Frenchman
> ___
> Therefore, Pierre likes red wine

The two statements above the line are called the *premises* of the inference, while the statement below the line is called the *conclusion*. This is a deductive inference because it has the following property: *if* the premises are true, then the conclusion must be true too. If it's true that all Frenchmen like red wine, and that Pierre is a Frenchman, it follows that Pierre does indeed like red wine. This is sometimes expressed by saying that the premises of the inference *entail* the conclusion. Of course the premises of this inference are almost certainly not true—there are bound to be Frenchmen who do not like red wine. But that is not the point. What makes the inference deductive is the existence of an appropriate relation between premises and conclusion, namely that the truth of the premises guarantees the truth of the conclusion.

Not all inferences are deductive. Consider the following example:

> The first five eggs in the box were good.
> All the eggs have the same best-before date stamped on them.
>
> ---
>
> Therefore, the sixth egg will be good too.

This looks like a perfectly sensible piece of reasoning. But nonetheless it is not deductive, for the premises do not entail the conclusion. Even if the first five eggs were good, and all the eggs do have the same date stamp, it is quite conceivable that the sixth egg will be rotten. That is, it is logically possible for the premises of this inference to be true and yet the conclusion false, so the inference is not deductive. Instead it is known as an *inductive* inference. In a typical inductive inference, we move from premises about objects that we have examined to conclusions about objects of the same sort that we haven't examined—in this example, eggs.

Deductive inference is safer than its inductive cousin. When we reason deductively, we can be certain that if we start with true premises we will end up with a true conclusion. By contrast, inductive reasoning is quite capable of taking us from true premises to a false conclusion. Despite this defect, we seem to rely on inductive reasoning throughout our lives. For example, when you turn on your computer in the morning, you are confident it will not explode in your face. Why? Because you turn on your computer every morning, and it has never exploded up to now. But the inference from 'up until now, my computer has not exploded when I turned it on' to 'my computer will not explode this time' is inductive, not deductive. It is logically possible that your computer will explode this time, even though it has never done so before.

Do scientists use inductive reasoning too? The answer seems to be yes. Consider the condition known as Down's syndrome (DS). Geneticists tell us that people with DS have three copies of chromosome 21 instead of the usual two. How do they know this? The answer, of course, is that they examined a large number of people with DS and found that each had an additional copy of chromosome 21. They then reasoned inductively to the conclusion that *all* people with DS, including those they hadn't examined, have an additional copy. This inference is inductive not deductive. For it is possible, though unlikely, that the sample examined was unrepresentative. This example is not an isolated one. In effect, scientists reason inductively whenever they move from limited data to a more general conclusion, which they do all the time.

The central role of induction in science is sometimes obscured by how we talk. For example, you might read a newspaper report which says that scientists have found 'experimental proof' that genetically modified maize is safe to eat. What this means is that the scientists have tested the maize on a large number of people and none have come to any harm. But strictly speaking this doesn't *prove* that the maize is safe, in the sense in which mathematicians can prove Pythagoras' theorem, say. For the inference from 'the maize didn't harm any of the people on whom it was tested' to 'the maize will not harm anyone' is inductive, not deductive. The newspaper report should really have said that scientists have found good *evidence* that the maize is safe for humans. The word 'proof' should strictly only be used when we are dealing with deductive inferences. In this strict sense of the word, scientific hypotheses can rarely if ever be proved true by the data.

Most philosophers think it's obvious that science relies heavily on induction, indeed so obvious that it hardly needs arguing for. But remarkably, this was denied by the philosopher Karl Popper, whom we met in the last chapter. Popper claimed that scientists only need to use deductive inferences. This would be nice if it were true, for deductive inferences are safer than inductive ones, as we have seen.

Popper's basic argument was this. Although a scientific theory (or hypothesis) can never be proved true by a finite amount of data, it can be proved false, or refuted. Suppose a scientist is testing the hypothesis that all pieces of metal conduct electricity. Even if every piece of metal they examine conducts electricity, this doesn't prove that the hypothesis is true, for reasons that we've seen. But if the scientist finds even one piece of metal that fails to conduct electricity, this conclusively refutes the theory. For the inference from 'this piece of metal does not conduct electricity' to 'it is false that all pieces of metal conduct electricity' is a deductive inference—the premise entails the conclusion. So if a scientist were trying to refute their theory, rather than establish its truth, their goal could be accomplished without the use of induction.

The weakness of Popper's argument is obvious. For the goal of science is not solely to refute theories, but also to determine which theories are true (or probably true). When a scientist collects experimental data, their aim *might* be to show that a particular theory—their arch-rival's theory perhaps—is false. But much more likely, they are trying to convince people that their own theory is true. And in order to do that, they will have to resort to inductive reasoning of some sort. So Popper's attempt to show that science can get by without induction does not succeed.

Hume's problem

Although inductive reasoning is not logically watertight, it seems like a sensible way of forming beliefs about the world. Surely the fact that the sun has risen every day in the past gives us good reason to believe that it will rise tomorrow? If you came across someone who professed to be entirely agnostic about whether the sun will rise tomorrow or not, you would regard them as very strange indeed, if not irrational.

But what justifies this faith we place in induction? How should we go about persuading someone who refuses to reason inductively that they are wrong? The 18th-century Scottish philosopher David Hume (1711–76) gave a simple but radical answer to this question. He argued that the use of induction cannot be rationally justified at all. Hume admitted that we use induction all the time, in everyday life and in science, but insisted that this was a matter of brute animal habit. If challenged to provide a good reason for using induction, we can give no satisfactory answer, he thought.

How did Hume arrive at this startling conclusion? He began by noting that whenever we make inductive inferences, we seem to presuppose what he called the 'uniformity of nature'. To see what Hume meant by this, recall our examples. We had the inference from 'the first five eggs in the box were good' to 'the sixth egg will be good'; from 'the Down's syndrome patients examined had an extra chromosome' to 'all those with Down's syndrome have an extra chromosome'; and from 'my computer has never exploded until now' to 'my computer will not explode today'. In each case, our reasoning seems to depend on the assumption that objects we haven't examined will be similar, in relevant respects, to objects of the same sort that we have examined. That assumption is what Hume means by the uniformity of nature.

But how do we know that the uniformity assumption is true? Can we perhaps prove its truth somehow? No, says Hume, we cannot. For it is easy to *imagine* a world where nature is not uniform but changes its course randomly from day to day. In such a world, computers might sometimes explode for no reason, water might sometimes intoxicate us without warning, and billiard balls might sometimes stop dead on colliding. Since such a non-uniform world is conceivable, it follows that we cannot prove that the uniformity assumption is true. For if we could, then the non-uniform universe would be a logical impossibility.

Even if we cannot prove the uniformity assumption, we might nonetheless hope to find good empirical evidence for its truth. After all, the assumption has always held good up to now, so surely this is evidence that it is true? But this begs the question, says Hume! Grant that nature has behaved largely uniformly up to now. We cannot appeal to this fact to argue that nature will continue to be uniform, says Hume, because this assumes that what has happened in the past is a reliable guide to what will happen in the future—which *is* the uniformity of nature assumption. If we try to argue for the uniformity assumption on empirical grounds, we end up reasoning in a circle.

The force of Hume's point can be appreciated by imagining how you would persuade someone who doesn't trust inductive reasoning that they should. You might say: 'look, inductive reasoning has worked pretty well up until now. By using induction scientists have split the atom, landed on the moon, and invented lasers. Whereas people who haven't used induction have died nasty deaths. They have eaten arsenic believing it would nourish them, and jumped off tall buildings believing they would fly. Therefore it will clearly pay you to reason inductively.' But this wouldn't convince the doubter. For to argue that induction is trustworthy because it has worked well up to now is to reason inductively. Such an argument would carry no weight with someone who doesn't *already* trust induction. That is Hume's fundamental point.

This intriguing argument has exerted a powerful influence on the philosophy of science. (Popper's attempt to show that science need only use deduction was motivated by his belief that Hume had shown the unjustifiability of induction.) The influence of Hume's argument is not hard to understand. For normally we think of science as the very paradigm of rational enquiry. We place great faith in what scientists tell us about the world. But science relies on induction, and Hume's argument seems to show that induction cannot be rationally justified. If Hume is right, the foundations on which science is built do not look as solid as we might have hoped. This puzzling state of affairs is known as *Hume's problem of induction*.

Philosophers have responded to Hume's problem in literally dozens of ways; this is still an active area of research today. One response says that to seek a 'justification of induction', or to bemoan the lack of one, is ultimately incoherent. Peter Strawson, an Oxford philosopher from the 1950s, defended this view with the following analogy. If someone worried whether a particular action was legal, they could consult the lawbooks and see what they say. But suppose someone worried about whether the law itself was legal. This is an odd worry indeed. For the law is the standard against which the legality of other things is judged, and it makes little sense to enquire whether the standard itself is legal. The same applies to induction, Strawson argued. Induction is one of the standards we use to decide whether someone's beliefs about the world are justified. So it makes little sense to ask whether induction itself is justified.

Has Strawson really succeeded in defusing Hume's problem? Some philosophers say yes, others say no. But most agree that it is very hard to see how there *could* be a satisfactory justification of induction. (Frank Ramsey, a famous Cambridge philosopher, wrote in 1919 that to ask for a justification of induction was 'to cry for the moon'.) Whether this is something that should worry us, or shake our faith in science, is a difficult question that you should ponder for yourself.

Inference to the best explanation

The inductive inferences we've examined so far have all had essentially the same structure. In each case, the premise has had the form 'all examined *Fs* have been *G*', and the conclusion the form 'other *Fs* are also *G*'. In short, these inferences take us from examined to unexamined instances of a given kind.

Such inferences are widely used in everyday life and in science, as we have seen. However, there is another common type of non-deductive inference which doesn't fit this simple pattern. Consider the following example:

> The cheese in the larder has disappeared, apart from a few crumbs.
> Scratching noises were heard coming from the larder last night.
>
> ---
>
> Therefore, the cheese was eaten by a mouse.

It is obvious that this inference is non-deductive: the premises do not entail the conclusion. For the cheese could have been stolen by the maid, who cleverly left a few crumbs to make it look like the handiwork of a mouse; and the scratching noises could have been caused by the boiler overheating. Nonetheless, the inference is clearly a reasonable one. For the hypothesis that a mouse ate the cheese seems to provide a *better explanation* of the data than the 'maid and boiler' hypothesis. After all, maids do not normally steal cheese, and modern boilers rarely overheat. Whereas mice do eat cheese when they get the chance, and do make scratching sounds. So although we cannot be certain that the mouse hypothesis is true, on balance it looks plausible.

Reasoning of this sort is known as 'inference to the best explanation', or IBE for short. Certain terminological confusions surround the relation between IBE and induction. Some philosophers describe IBE as a *type* of inductive inference; in effect, they use 'inductive inference' to mean 'any inference which is not deductive'. Others *contrast* IBE with induction, as we have done. On this way of cutting the pie, 'induction' is reserved for inferences from examined to unexamined instances of a given kind; IBE and induction are then two different types of non-deductive inference. Nothing hangs on which choice of terminology we favour, so long as we stick to it consistently.

Scientists frequently use IBE. For example, Darwin argued for his theory of evolution by calling attention to various facts about the living world which are hard to explain if we assume that current species have been separately created, but which make perfect sense if current species have descended from common ancestors, as his theory held. For example, there are close anatomical similarities between the legs of horses and zebras. How do we explain this, if God created horses and zebras separately? Presumably he could have made their legs as different as he pleased. But if horses and zebras have descended from a common ancestor, this provides an obvious explanation of their anatomical similarity. Darwin argued that the ability of his theory to explain such facts constituted strong evidence for its truth. 'It can hardly be supposed', he wrote, 'that a false theory would explain, in so satisfactory a manner as does the theory of natural selection, the several large classes of fact above specified.'

Another example of IBE is Einstein's famous work on Brownian motion—the zig-zag motion of microscopic particles suspended in a liquid or gas. A number of attempted explanations of Brownian motion were advanced in the 19th century. One theory attributed the motion to electrical attraction between particles, another to agitation from external surroundings, and another to convection currents in the fluid. The correct explanation is based on the kinetic theory of matter, which says that liquids and gases are made up of atoms or molecules in motion. The suspended particles collide with the surrounding molecules, causing their erratic movements. This theory was proposed in the late 19th century but not widely accepted, not least because many scientists didn't believe that atoms and molecules were real entities. But in 1905, Einstein provided an ingenious mathematical treatment of Brownian motion, making a number of predictions that were later confirmed experimentally. After Einstein's work, the kinetic theory was quickly agreed to provide a better explanation of Brownian motion than the alternatives, and scepticism about the existence of atoms and molecules subsided.

The basic idea behind IBE—reasoning from one's data to a theory or hypothesis that explains the data—is straightforward. But how do we decide which of the competing hypotheses provides the 'best explanation' of the data? What criteria determine this? One popular answer is that a good explanation should be simple, or parsimonious. Consider again the cheese-in-the-larder example. There are two pieces of data that need explaining: the missing cheese and the scratching noises. The mouse hypothesis postulates just one cause—a mouse—to explain both pieces of data. But the maid-and-boiler hypothesis must postulate two causes—a dishonest maid and an overheating boiler—to explain the same data. So the mouse hypothesis is more parsimonious, hence better. The Darwin example is similar. Darwin's theory could explain a diverse range of facts about the living world, not just anatomical similarities between species. Each of these facts could in principle be explained in other ways, but the theory of evolution explained all the facts in one go—that is what made it the best explanation of the data.

The idea that simplicity or parsimony is the mark of a good explanation is quite appealing, and helps flesh out the abstract idea of IBE. But if scientists use simplicity as a guide to inference, this raises a deep question. Do we have reason to think that the universe is simple rather than complex? Preferring a theory which explains the data in terms of the fewest number of causes seems sensible. But are there any objective grounds for thinking that such a theory is more likely to be true than a less simple rival? Or is simplicity something that scientists value because it makes their theories easier to formulate and to understand? Philosophers of science do not agree on the answer to this difficult question.

Causal inference

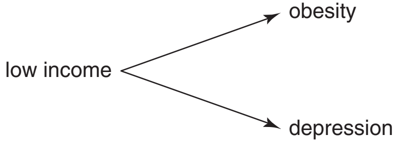
A key goal of science is to discover the causes of natural phenomena. Often this quest is successful. For example, climate change scientists know that burning fossil fuels causes global warming; chemists know that heating a liquid causes it to become a gas; and epidemiologists know that the MMR vaccine does not cause autism. Since causal connections are not directly observable (as David Hume famously argued), scientific knowledge of this sort must be the result of inference. But how exactly does causal inference work?

It is helpful to distinguish two cases: inferring the cause of a particular event versus inferring a general causal principle. To illustrate the distinction, consider the contrast between ‘a meteorite strike caused the extinction of the dinosaurs’ and ‘smoking causes lung cancer’. The former is a singular statement about the cause of a particular historical event, the latter a general statement about the cause of a certain *sort* of event (getting lung cancer). In both cases a process of inference has led scientists to believe the statements in question, but the inferences work in somewhat different ways. Here we focus on inferences of the second sort, i.e. to general causal principles.

Suppose a medical researcher wishes to test the hypothesis that obesity causes depression. How should they proceed? A natural first step is to see whether the two attributes are correlated. To assess this, they could examine a large sample of obese people, and see whether the incidence of depression is higher in this group than in the general population. If it is, then unless there is some reason to think the sample unrepresentative, it is reasonable to infer (by ordinary induction) that obesity and depression are correlated in the overall population.

Would such a correlation show that obesity causes depression? Not necessarily. First-year science students are routinely taught that correlation does not imply causation, and with good reason. For there are other possible explanations of the correlation. The direction of causation could be the other way round, i.e. being depressed might cause people to eat more, hence to become obese. Or there might be no causal influence of obesity on depression nor vice versa, but the two conditions are joint effects of a common cause. For example, perhaps low income raises the chance of obesity and also raises the chance of depression via a separate causal pathway (see Figure 3). If so, we would expect obesity and depression to be correlated in the population. This 'common cause' scenario is a major reason why causation cannot always be reliably inferred from correlational data.

How could we test the hypothesis that low income causes both obesity and depression? The obvious thing to do is to find a sample of individuals *all with the same income level*, and examine



3. Causal graph depicting the hypothesis that low income is a common cause of both obesity and depression.

whether obesity and depression are correlated within the sample. If we do this for a number of different income levels, and find that within each income-homogeneous sample the correlation disappears, this is strong evidence in favour of the common cause hypothesis. For it shows that once income is taken into account, obesity is no longer associated with depression. Conversely, if a strong obesity–depression correlation exists even among individuals with the same income level, this is evidence against the common cause hypothesis. In statistical jargon, this procedure is known as 'controlling for' the variable income.

The underlying logic here is similar to that of the controlled experiment, a mainstay of modern science. Suppose an entomologist wishes to test the hypothesis that rearing insect larvae at higher temperatures leads to reduced adult body size. To test this, the entomologist gets a large number of insect larvae, rears some at a cool and others at a warm temperature, then measures the size of the resulting adults. For this to be an effective test of the causal hypothesis, it is important that all factors other than temperature be held constant between the two groups, so far as possible. For example, the larvae should all be from the same species, the same sex, and be fed the same food. So the entomologist must design their experiment carefully, controlling for all variables that could potentially affect adult body size. Only then can a difference in adult body size between the two groups safely be attributed to the temperature difference.

It is sometimes argued that controlled experiments are the only reliable way of making causal inferences in science. Proponents of this view argue that purely observational data, without any experimental intervention, cannot give us knowledge of causality. However this is a controversial thesis. For while controlled experimentation is certainly a good way of probing nature's secrets, the technique of statistical control can often accomplish something quite similar. In recent years, statisticians and computer scientists have developed powerful techniques for making causal inferences from observational data. Whether there is a fundamental methodological difference between experimental and observational data, vis-à-vis the reliability of the causal inferences that can be drawn from them, is a matter of continuing debate.

In modern biomedical science, a particular sort of controlled experiment is often given particular prominence. This is the *randomized controlled trial* (RCT), originally devised by R. A. Fisher in the 1930s, and often used to test the effectiveness of a new drug. In a typical RCT, patients with a particular medical condition, e.g. severe migraine, are divided into two groups. Those in the treatment group receive the drug, while those in the control group do not. The researchers then compare the two groups on the outcome of interest, e.g. relief of migraine symptoms. If those in the treatment group do significantly better than the control group, this is presumptive evidence that the drug works. The key feature of an RCT is that the initial division of the patients into two groups must be done at *random*. Fisher and his modern followers argue that this is necessary to sustain a valid causal inference.

Why is randomization so important? Because it helps to eliminate the effect of confounding factors on the outcome of interest. Typically the outcome will be affected by many factors, e.g. age, diet, and exercise. Unless all of these factors are known, the researcher cannot explicitly control for them. However by randomly allocating patients to the treatment and control groups, this problem can be largely circumvented. Even if factors other than the drug do affect the outcome, randomization ensures that any such factors are unlikely to be over-represented in either the treatment or the control group. So if there is a significant difference in outcome between treatment and control groups, this is very likely due to the drug. Of course this does not strictly *prove* that the drug was causally responsible, but it constitutes strong evidence.

In medicine, the RCT is usually regarded as the 'gold standard' for assessing causality. Indeed proponents of the movement known as 'evidence-based medicine' often argue that *only* an RCT can tell us when a particular treatment is causally effective. However this position is arguably too strong (and the appropriation of the word 'evidence' to refer only to RCTs is misleading). In many areas of science, RCTs are not feasible, for either practical or ethical reasons, and yet causal inferences are routinely made. Furthermore, much of the causal knowledge we have in everyday life we gained without RCTs. Young children know that putting their hand in the fire causes a painful burning sensation; no randomized trial was needed to establish this. While RCTs are certainly important, and should be done when feasible, it is not true that they are the only way of discovering causality.

Probability and scientific inference

Given that inductive reasoning cannot give us certainty, it is natural to hope that the concept of probability will help us understand how it works. Even if a scientist's evidence does not prove that their hypothesis is true, surely it can render it highly probable? Before exploring this idea we need to attend briefly to the concept of probability itself.

Probability has both an objective and a subjective guise. In its objective guise, probability refers to how often things in the world happen, or tend to happen. For example, if you are told that the probability of an Englishwoman living to age 90 is one in ten, you would understand this as meaning that one-tenth of all Englishwomen attain that age. Similarly, a natural understanding of the statement 'the probability that the coin will land heads is a half' is that in a long sequence of coin flips, the proportion of heads would be very close to a half. Understood this way, statements about probability are objectively true or false, independently of what anyone believes.

In its subjective guise, probability is a measure of rational degree of belief. Suppose a scientist tells you that the probability of finding life on Mars is extremely low. Does this mean that life is found on only a small proportion of all the celestial bodies? Surely not. For one thing, no one knows how many celestial bodies there are, nor how many of them contain life. So a different notion of probability is at work here. Now since there either is life on Mars or there isn't, talk of probability in this context must presumably reflect our ignorance of the state of the world, rather than describing an objective feature of the world itself. So it is natural to take the scientist's statement to mean that in the light of all the evidence, the rational degree of belief to have in the hypothesis that there is life on Mars is very low.

The idea that the rational degree of belief to have in a scientific hypothesis, given the evidence, may be viewed as a type of probability suggests a natural picture of how scientific inference works. Suppose a scientist is considering a particular hypothesis H. In the light of the evidence to date, the scientist has a certain degree of belief in H, denoted $P(H)$, which is a number between zero and one. (Another name for $P(H)$ is the scientist's 'credence' in H.) Some new evidence then comes to light, e.g. from experiment or observation. In the light of this new evidence, the scientist updates their credence in H to $P_{new}(H)$. If the new evidence supports the theory, then $P_{new}(H)$ will be greater than $P(H)$, i.e. the scientist will have become more confident that H is true.

A toy example will help flesh this out. Suppose a playing card has been drawn from a well-shuffled pack and is concealed from your view. Let H be the hypothesis that the card is the queen of hearts. What is the value of $P(H)$, i.e. your initial rational credence in H? Presumably it is 1/52. For there are fifty-two cards in the pack and they are all equally likely to be chosen. Suppose you then learn that the chosen card is definitely a heart. Call this piece of information e. In the light of e, what is the value of $P_{new}(H)$, i.e. your updated credence in H given the new evidence? Clearly, $P_{new}(H)$ should equal 1/13—for there are thirteen hearts in the pack and you know that the concealed card is one of them. So learning e has increased your credence in H from 1/52 to 1/13.

This is all fairly obvious, but what is the general rule for updating your credence in the light of new information? The answer is called 'conditionalization'. To grasp this rule we need the concept of conditional probability. In the card example, $P(H)$ is your initial credence in hypothesis H. Your initial credence in H conditional on the assumption that e is true is denoted $P(H/e)$. (Read this as 'the probability of H given e'.) What is the value of $P(H/e)$? The answer is 1/13. For on the assumption that e is true, i.e. that the card drawn is a heart, your credence in the hypothesis H equals 1/13. When you learn that e *is* actually true, your new credence in H, i.e. $P_{new}(H)$, should then be set equal to your initial credence in H conditional on e, according to the rule of conditionalization.

Rule of conditionalization

Upon learning evidence e, $P_{new}(H)$ should equal $P(H/e)$.

To better understand the rule of conditionalization, note that the conditional probability $P(H/e)$ is by definition equal to the ratio $P(H \text{ and } e)/P(e)$. In the card example, $P(H \text{ and } e)$ denotes your initial credence that both H and e are true. But since in this case H logically entails e—for if the card is the queen of hearts then it must be a heart—it follows that $P(H \text{ and } e)$ is simply equal to $P(H)$, i.e. 1/52. What about $P(e)$? This is your initial credence that the chosen card is a heart. Since exactly one quarter of the cards in the deck are hearts, and you regard all the cards as equally likely to be the chosen one, it follows that $P(e)$ is ¼. Applying the definition of $P(H/e)$, this tells us that $P(H/e)$ equals 1/52 divided by ¼, which is 1/13—the same answer as we computed previously.

The rule of conditionalization may sound complicated, but like many logical rules we often obey it without thinking. In the card example, it is intuitively obvious that learning e should increase your rational credence in H from 1/52 to 1/13, and in practice this is exactly what most people would do. In doing so, they are implicitly obeying the rule of conditionalization even if they have never heard of it. In addition to its implicit uses, the conditionalization rule is often used explicitly by scientists, for example in certain sorts of statistical reasoning. The branch of statistics known as Bayesian statistics makes extensive use of updating by conditionalization. (The name 'Bayesian' refers to the 17th-century English clergyman Thomas Bayes, an early pioneer of probability theory, who discovered the conditionalization rule.)

Some philosophers of science wish to use updating by conditionalization as a general model for scientific inference, applicable even to inferences that are not explicitly probabilistic. The idea is that any rational scientist can be thought of as having an initial credence in their theory or hypothesis, which they then update in the light of new evidence by following the rule of conditionalization. Even if the scientist's conscious reasoning process looks nothing like this, it is a useful idealization according to these philosophers.

This 'Bayesian' view of scientific inference is quite attractive, as it sheds light on certain aspects of the scientific method. Consider the fact that when a scientific theory makes a testable prediction that turns out to be true, this is usually taken as evidence in favour of the theory. In Chapter 1 we had the example of Einstein's theory of general relativity predicting that starlight would be deflected by the sun's gravitational field; when this prediction was confirmed it increased scientists' confidence in Einstein's theory. But why should a successful prediction enhance a scientist's confidence in a theory, given that there will always be other possible explanations that can't be ruled out? Is this simply a brute fact about how scientists reason, or does it have a deeper explanation?

Bayesians argue that it does indeed have a deeper explanation. Suppose that a theory T entails a testable statement e. The scientist initially has credence $P(T)$ that T is true and $P(e)$ that e is true. We assume that both $P(T)$ and $P(e)$ take non-extreme values, i.e. are not zero or one. Suppose the scientist then learns that e is definitely true. If they follow the rule of conditionalization, their new credence in theory T, i.e. $P_{new}(T)$, must then be greater than $P(T)$ as a matter of logic. In other words, upon learning that their theory has made a true prediction, a scientist will necessarily increase their confidence in the theory so long as they obey the conditionalization rule. So the fact that successful predictions typically lead scientists to become more confident of their theories has a neat explanation, on the Bayesian view of scientific inference.

However the Bayesian view has its limitations. Much interesting scientific inference involves inventing theories or hypotheses that have never been thought of before. The great scientific advances made by Copernicus, Newton, and Darwin were all of this sort. Each of these scientists came up with a new theory which their predecessors had never entertained. The reasoning that led them to these theories cannot plausibly be regarded as Bayesian. For conditionalization describes how a scientist's rational credence in a theory should change when they get new evidence; this presumes that the theory has already been thought of. So scientific inferences that go from data to completely new theory cannot be understood in terms of conditionalization.

Another limitation of the Bayesian view concerns the source of the initial credences, prior to updating on the new evidence. In the card example, your initial rational credence that the chosen card was the queen of hearts was easy to determine, because there are fifty-two cards in a deck each with an equal chance of being chosen. But many scientific hypotheses are not like this. Consider the hypothesis that global warming will exceed four degrees by the year 2100. What should a scientist's initial credence in this hypothesis, before getting any relevant evidence, be? There is no obvious answer to this question. Some Bayesian philosophers of science reply that initial credences are purely subjective, i.e. they simply represent a scientist's 'best guess' about the hypothesis, so any initial credence is as good as any other. On this version of the Bayesian view, there is an objectively rational way for a scientist to *change* their credences when they get new evidence, i.e. conditionalization, but no objective constraint on what their initial credences should be.

This intrusion of a subjective dimension is regarded as unwelcome by many philosophers, leading them to conclude that the Bayesian view cannot be the whole story about scientific inference. Also, it shows that there cannot be a Bayesian 'solution' to Hume's problem of induction. The idea that we can somehow escape Hume's problem by invoking probability is an old one. Even if the sun's having risen every day in the past doesn't prove that it will rise tomorrow, surely it makes it highly probable? Whether this response to Hume ultimately works is a complex matter, but we can say the following. If the only objective constraints concern how we should change our credences, but what our initial credences should be is entirely subjective, then individuals with very bizarre opinions about the world will count as perfectly rational. So a probabilistic escape from Hume's problem will not fall out of the Bayesian view of scientific inference.

Chapter 3
Explanation in science

One important aim of science is to try and explain what happens in the world around us. Sometimes we seek explanations for practical ends. For example, we might want to know why the ozone layer is being depleted so quickly in order to try and do something about it. In other cases we seek scientific explanations simply to satisfy our intellectual curiosity—we want to understand more about how the world works. Historically, the pursuit of scientific explanation has been motivated by both goals.

Quite often, modern science is successful in its aim of supplying explanations. For example, chemists can explain why sodium turns yellow when it burns. Astronomers can explain why solar eclipses occur when they do. Economists can explain why the yen declined in value in the 1980s. Geneticists can explain why male baldness tends to run in families. Neurophysiologists can explain why extreme oxygen deprivation leads to brain damage. You can probably think of many other examples of successful scientific explanations.

But what exactly *is* a scientific explanation? What exactly does it mean to say that a phenomenon can be 'explained' by science? This is a question that has exercised philosophers since Aristotle, but our starting-point will be a famous account of scientific explanation put forward in the 1950s by the German-American philosopher

Carl Hempel. Hempel's account is known as the *covering law* model of explanation, for reasons that will become clear.

Hempel's covering law model of explanation

The basic idea behind the covering law model is straightforward. Hempel noted that scientific explanations are usually given in response to what he called 'explanation-seeking why-questions'. These are questions such as 'why is the earth not perfectly spherical?' or 'why do women live longer than men?'—they are demands for explanation. To give a scientific explanation is thus to provide a satisfactory answer to an explanation-seeking why-question. If we could determine the essential features that such an answer must have, we would know what scientific explanation is.

Hempel suggested that scientific explanations typically have the logical structure of an *argument*, i.e. a set of premises followed by a conclusion. The conclusion states that the phenomenon which needs explaining occurs, and the premises tell us why the conclusion is true. Thus suppose someone asks why sugar dissolves in water. This is an explanation-seeking why-question. To answer it, says Hempel, we must construct an argument whose conclusion is 'sugar dissolves in water' and whose premises tell us why this conclusion is true. The task of providing an account of scientific explanation then becomes the task of characterizing precisely the relation that must hold between a set of premises and a conclusion, in order for the former to count as an explanation of the latter. That was the problem Hempel set himself.

Hempel's answer to the problem was threefold. First, the premises should entail the conclusion, i.e. the argument should be a deductive one. Secondly, the premises should all be true. Thirdly, the premises should consist of at least one general law. General laws are things such as 'all metals conduct electricity', 'a body's acceleration varies inversely with its mass', and 'all plants contain chlorophyll'; they contrast with particular facts such as 'this piece of metal conducts electricity' and 'the plant on my desk contains chlorophyll'. General laws are sometimes called *laws of nature*. Hempel allowed that a scientific explanation could appeal to particular facts as well as general laws, but he held that at least one general law was always essential. So to explain a phenomenon, on Hempel's conception, is to show that its occurrence follows deductively from a general law, perhaps supplemented by other laws and/or particular facts, all of which must be true.

To illustrate, suppose I am trying to explain why the plant on my desk has died. I might offer the following explanation. Owing to the poor light in my study, no sunlight has been reaching the plant; but sunlight is necessary for a plant to photosynthesize; and without photosynthesis a plant cannot make the carbohydrates it needs to survive, and so will die; therefore my plant died. This explanation fits Hempel's model exactly. It explains the death of the plant by deducing it from two true laws—that sunlight is necessary for photosynthesis, and that photosynthesis is necessary for survival—and one particular fact—that the plant was not getting any sunlight. Given the truth of the two laws and the particular fact, the death of the plant *had* to occur; that is why the former constitute a good explanation of the latter.

Schematically, Hempel's model of explanation can be written as follows:

General Laws
Particular Facts
⇒
Phenomenon to be explained

The phenomenon to be explained is called the *explanandum*, and the general laws and particular facts that do the explaining are called the *explanans*. The *explanandum* may be either particular or general. In the previous example, it was a particular fact—the death of my plant. But sometimes the things we want to explain are general. For example, we might wish to explain why exposure to the sun often leads to skin cancer. This is itself a generality, not a particular fact. To explain it, we would need to deduce it from more fundamental laws—presumably, laws about the impact of radiation on skin cells, combined with particular facts about the amount of radiation in sunlight. So the structure of a scientific explanation is essentially the same whether the *explanandum*, i.e. thing we are trying to explain, is particular or general.

It is easy to see where the covering law model gets its name. For according to the model, the essence of explanation is to show that the phenomenon to be explained is 'covered' by some general law of nature. There is certainly something appealing about this idea. For showing that a phenomenon is a consequence of a general law takes the mystery out of it—it renders it more intelligible. And many actual scientific explanations do fit the pattern Hempel describes. For example, Newton explained why the planets move in ellipses around the sun by showing that this can be deduced from his law of universal gravitation, along with some minor additional assumptions. Newton's explanation fits Hempel's model exactly: a phenomenon is explained by showing that it had to be so, given the laws of nature plus some additional facts. After Newton, there was no longer any mystery about why planetary orbits are elliptical.

Hempel was aware that not all scientific explanations fit his model exactly. For example, if you ask someone why the smog in Athens has worsened in recent years they might reply 'because of the increase in domestic wood-burning'. This is true, and is a perfectly acceptable scientific explanation, though it involves no mention of any laws. But Hempel would say that if the explanation were spelled out in full detail, laws would enter the picture. Presumably there is a law which says something like 'if wood-smoke emissions exceed a certain level in an area of a given size, and if the wind is sufficiently light, smog clouds will form'. The full explanation of why the smog in Athens has worsened would cite this law, along with the fact that wood-burning in Athens has increased and that wind levels there are fairly low. In practice we wouldn't spell out the explanation in this much detail unless we were being very pedantic. But if we were to spell it out, it would correspond quite well to the covering law pattern.

Hempel drew an interesting consequence from his model about the relation between explanation and prediction. He argued that these are two sides of the same coin. Whenever we give a covering law explanation of a phenomenon, the laws and particular facts we cite would have enabled us to predict the occurrence of the phenomenon, if we hadn't already known about it. To illustrate, consider again Newton's explanation of why planetary orbits are elliptical. This fact was known long before Newton explained it using his theory of gravity—it was discovered by Kepler. But had it not been known, Newton would have been able to predict it from his theory of gravity. Hempel expressed this by saying that every scientific explanation is potentially a prediction—it would have served to predict the phenomenon in question, had it not already been known. The converse is also true, Hempel thought: every reliable prediction is potentially an explanation. To illustrate, suppose scientists predict that mountain gorillas will be extinct by 2030, based on information about the destruction of their habitat. Suppose they turn out to be right. According to Hempel, the information they used to predict the gorillas' extinction before it happened will serve to explain that same fact after it has happened. Explanation and prediction are structurally symmetric.

Though the covering law model captures the structure of many actual scientific explanations quite well, it also faces a number of awkward counterexamples. In particular, there are cases that fit the covering law model but intuitively do not count as genuine scientific explanations. These cases suggest that Hempel's model is too liberal—it allows in things that should be excluded. We focus on two such cases here.

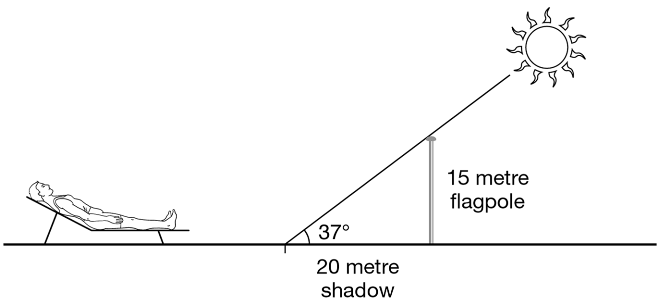
Case (i): the problem of symmetry

Suppose you are lying on the beach on a sunny day, and you notice that a flagpole is casting a shadow of 20 metres across the sand (see Figure 4).

Someone asks you to explain why the shadow is 20 metres long. This is an explanation-seeking why-question. A plausible answer might go as follows: 'light rays from the sun are hitting the flagpole, which is exactly 15 metres high. The angle of elevation of the sun is 37°. Since light travels in straight lines, a simple trigonometric calculation (tan 37° = 15/20) shows that the flagpole will cast a shadow 20 metres long.'

This looks like a perfectly good scientific explanation. And by rewriting it in accordance with Hempel's schema, we can see that it fits the covering law model:

General laws	Light travels in straight lines
	Laws of trigonometry
Particular facts	Angle of elevation of sun is 37°
	Flagpole is 15 metres high
Phenomenon to be explained	Shadow is 20 metres long



4. A 15-metre flagpole casts a shadow of 20 metres when the sun is 37° overhead.

The length of the shadow is deduced from the height of the flagpole and the angle of elevation of the sun, along with the optical law that light travels in straight lines and the laws of trigonometry. Since these laws are true, and since the flagpole is indeed 15 metres high, the explanation satisfies Hempel's requirements precisely. So far so good. The problem arises as follows. Suppose we swap the *explanandum*—that the shadow is 20 metres long—with the particular fact that the flagpole is 15 metres high. The result is this:

General laws	Light travels in straight lines
	Laws of trigonometry
Particular facts	Angle of elevation of sun is 37°
	Shadow is 20 metres long
Phenomenon to be explained	Flagpole is 15 metres high

This 'explanation' clearly conforms to the covering law pattern too. The height of the flagpole is deduced from the length of the shadow it casts and the angle of elevation of the sun, along with the optical law that light travels in straight lines and the laws of trigonometry. But it seems very odd to regard this as an *explanation* of why the flagpole is 15 metres high. The real explanation of why the flagpole is 15 metres high is presumably that a carpenter deliberately made it so—it has nothing to do with the length of the shadow that it casts. So Hempel's model is too liberal: it allows something to count as a scientific explanation which obviously is not.

The general moral of the flagpole example is that the concept of explanation exhibits an important asymmetry. The height of the flagpole explains the length of the shadow, given the relevant laws and additional facts, but not vice versa. In general, if x explains y, given the relevant laws and additional facts, then it will not be true that y explains x, given the same laws and facts. This is sometimes expressed by saying that explanation is an asymmetric relation. Hempel's covering law model does not respect this asymmetry. For just as we can deduce the length of the shadow from the height of the flagpole, given the laws and additional facts, so we can deduce the height of the flagpole from the length of the shadow. So Hempel's model fails to capture fully what it is to be a scientific explanation, for it implies that explanation should be a symmetric relation when in fact it is asymmetric.

The shadow and flagpole case also provides a counterexample to Hempel's thesis that explanation and prediction are two sides of the same coin. The reason is obvious. Suppose you didn't know how high the flagpole was. If someone told you that it was casting a shadow of 20 metres and that the sun was 37° overhead, you would be able to *predict* the flagpole's height, given that you knew the relevant optical and trigonometrical laws. But as we have just seen, this information clearly doesn't *explain* why the flagpole has the height it does. So in this example prediction and explanation part ways. Information that serves to predict a fact before we know it does not serve to explain that same fact after we know it, which contradicts Hempel's thesis.

Case (ii): the problem of irrelevance

Suppose a young child is in a maternity ward in a hospital. The child notices that one person in the room—who is a man called John—is not pregnant, and asks the doctor why not. The doctor replies: 'John has been taking birth control pills regularly for the last few years. People who take birth control pills regularly never become pregnant. Therefore, John has not become pregnant.' Let us suppose that what the doctor says is true—John is mentally ill and does indeed take birth control pills, which he believes help him. Even so, the doctor's reply to the child is clearly not helpful. The correct explanation of why John has not become pregnant, obviously, is that he is male and males cannot become pregnant.

However, the explanation the doctor has given fits the covering law model exactly. The doctor deduces the phenomenon to be explained—that John is not pregnant—from the general law that people who take birth control pills do not become pregnant and the particular fact that John has been taking birth control pills. Since both the general law and the particular fact are true, and since they do entail the *explanandum*, according to the covering law model the doctor has given an explanation of why John is not pregnant. But of course he hasn't.

The general moral is that a good explanation of a phenomenon should contain information that is *relevant* to the phenomenon's occurrence. This is where the doctor's reply to the child goes wrong. Although what the doctor tells the child is perfectly true, the fact that John has been taking birth control pills is irrelevant to his not being pregnant, because he wouldn't have been pregnant even if he hadn't taken been taking the pills. This is why the doctor's reply does not constitute a good answer to the child's question. Hempel's model does not respect this crucial feature of our concept of explanation.

Explanation and causality

Since the covering law model encounters problems, it is natural to look for an alternative way of understanding scientific explanation. Some philosophers believe that the key lies in the concept of *causality*. This is quite an attractive suggestion. For in many cases to explain a phenomenon is indeed to say what caused it. For example, if an accident investigator is trying to explain an aeroplane crash, they are obviously looking for the cause of the crash. Indeed the questions 'why did the plane crash?' and 'what was the cause of the plane crash?' are practically synonymous. Similarly, if an ecologist is trying to explain why there is less biodiversity in the tropical rainforests than there used to be, they are looking for the cause of the reduction in biodiversity. The link between the concepts of explanation and causality is quite intimate.

Impressed by this link, a number of philosophers have abandoned the covering law account of explanation in favour of causality-based accounts. The details vary, but the basic idea behind these accounts is that to explain a phenomenon is simply to say what caused it. In some cases, the difference between the covering law and causal accounts is not actually very great, for to deduce the occurrence of a phenomenon from a general law often just *is* to give its cause. For example, recall again Newton's explanation of why planetary orbits are elliptical. We saw that this explanation fits the covering law model—for Newton deduced the shape of the planetary orbits from his law of gravity, plus some additional facts. But Newton's explanation was also a causal one, since elliptical planetary orbits are *caused* by the gravitational attraction between planets and the sun.

However the covering law and causal accounts are not fully equivalent—in some cases they diverge. Indeed, many philosophers favour a causal account of explanation precisely because they think it can avoid some of the problems facing the covering law model. Recall the flagpole problem. Why do our intuitions tell us that the height of the flagpole explains the length of the shadow, given the laws, but not vice versa? Plausibly, because the height of the flagpole is the *cause* of the shadow being 20 metres long, but the shadow being 20 metres long is not the cause of the flagpole being 15 metres high. So unlike the covering law model, a causal account of explanation gives the 'right' answer in the flagpole case—it respects our intuition that we cannot explain the height of the flagpole by pointing to the length of the shadow it casts.

The general moral of the flagpole problem was that the covering law model cannot accommodate the fact that explanation is an asymmetric relation. Now causality is obviously an asymmetric relation too: if x is the cause of y, then y is not the cause of x. For example, if the short-circuit caused the fire, then the fire clearly did not cause the short-circuit. It is therefore natural to suggest that the asymmetry of explanation derives from the asymmetry of causality. If to explain a phenomenon is to say what caused it, then since causality is asymmetric we should expect explanation to be asymmetric too—as it is. The covering law model runs up against the flagpole problem precisely because it tries to analyse the concept of scientific explanation without reference to causality.

The same is true of the birth control pill case. That John takes birth control pills does not explain why he isn't pregnant, because the birth control pills are not the *cause* of his not being pregnant. Rather, John's sex is the cause of his not being pregnant. That is why we think that the correct answer to the question 'why is John not pregnant?' is 'because he is male, and males can't become pregnant', rather than the doctor's answer. So the covering law model runs into the problem of irrelevance precisely because it does not explicitly require that a scientific explanation identify the cause of the phenomenon that we wish to explain.

It is easy to criticize Hempel for failing to respect the close link between causality and explanation, as many philosophers have done. In some ways this criticism is a bit unfair. For Hempel subscribed to the philosophical doctrine called *empiricism*, and empiricists are traditionally suspicious of the concept of causality. Empiricism says that all our knowledge comes from experience. David Hume, whom we met in Chapter 2, was a leading empiricist, and he argued that it is impossible to experience causal relations. So he concluded that they don't exist—causality is something that we humans 'project' onto the world! This is a very hard conclusion to accept. Surely it is an objective fact that dropping glass vases causes them to break? Hume denied this. He allowed that it is an objective fact that most glass vases which have been dropped have in fact broken. But our idea of causality includes more than this. It includes the idea of a causal connection between the dropping and the breaking, i.e. that the former *brings about* the latter. No such connections are to be found in the world, according to

Hume: all we see is a vase being dropped, and then it breaking a moment later. This leads us to believe there is a causal connection between the two, but in reality there is not.

Few empiricists have accepted this startling conclusion outright. But as a result of Hume's work, they have tended to regard causality as a concept to be treated with caution. So to an empiricist, the idea of analysing explanation in terms of causality would seem perverse. If one's goal is to clarify the concept of scientific explanation, as Hempel's was, there is little point in using notions which are equally in need of clarification themselves. So the fact that the covering law model makes no mention of causality was not a mere oversight on Hempel's part. In recent years empiricism has declined somewhat in popularity. Furthermore, many philosophers have come to the conclusion that the concept of causality, although problematic, is indispensable to how we understand the world. So the idea of a causality-based account of scientific explanation seems more acceptable than it would have done in Hempel's day.

Causality-based accounts capture the structure of many actual scientific explanations quite well, but there are also cases they fit less well. Consider what are called 'theoretical identifications' in science, such as 'water is H_2O' or 'temperature is mean molecular kinetic energy'. In both cases, a familiar everyday concept is equated or identified with a more esoteric scientific concept. Such theoretical identifications furnish us with what appear to be scientific explanations. When chemists discovered that water is H_2O, they thereby explained what water is. Similarly, when physicists discovered that an object's temperature is the average kinetic energy of its molecules, they thereby explained what temperature is. But neither of these explanations is causal. Being made of H_2O doesn't *cause* a substance to be water—it just *is* being water. Having a particular mean molecular kinetic energy doesn't *cause* a liquid to have the temperature it does—it just *is* having that temperature. If these examples are accepted as legitimate scientific explanations, they suggest that causality-based accounts of explanation cannot be the whole story.

Can science explain everything?

Modern science can explain a great deal about the world we live in. But there are also numerous facts that have not been explained by science, or at least not explained fully. The origin of life is one such example. We know that about four billion years ago, molecules with the ability to make copies of themselves appeared in the primeval soup, and life evolved from there. But we do not understand how these self-replicating molecules got there in the first place (though some possible scenarios have been sketched). Another example is the fact that children with Asperger's syndrome often have very good memories. Numerous studies have confirmed this fact, but as yet nobody has succeeded in explaining it.

Many people believe that in the end, science will be able to explain facts of this sort. This is quite a plausible view. Molecular biologists are working hard on the problem of the origin of life, and only a pessimist would say they will never solve it. Admittedly the problem is not easy, not least because it is hard to know what conditions on earth four billion years ago were like. But nonetheless, there is no reason to think that the origin of life will never be explained. Similarly for the exceptional memories of children with Asperger's. The science of memory is still fairly new, and much remains to be discovered about the neurological basis of conditions such as Asperger's syndrome. Obviously we cannot guarantee that the explanation will eventually be found. But given the number of explanatory successes that modern science has already notched up, the smart money must be on many of today's unexplained facts eventually being explained too.

But does this mean that science can in principle explain everything? Or are there some phenomena that must forever elude scientific explanation? This is not an easy question to answer. On the one hand, it seems arrogant to assert that science can explain everything. On the other hand, it seems short-sighted to assert that any particular phenomenon can never be explained scientifically. For science changes and develops fast, and a phenomenon that looks completely inexplicable from the vantage-point of today's science may be easily explained tomorrow.

According to many philosophers, there is a purely logical reason why science will never be able to explain everything. For in order to explain something, whatever it is, we need to invoke something else. But what explains the second thing? To illustrate, recall that Newton explained a diverse range of phenomena using his law of gravity. But what explains the law of gravity itself? If someone asks *why* all bodies exert a gravitational attraction on each other, what should we tell them? Newton had no answer to this question. In Newtonian science the law of gravity was a fundamental principle: it explained other things, but could not itself be explained. The moral generalizes. However much the science of the future can explain, the explanations it gives will have to make use of certain fundamental laws and principles. Since nothing can explain itself, it follows that at least some of these laws and principles will themselves remain unexplained.

Whatever one makes of this argument, it is undeniably very abstract. It purports to show that some things will never be explained, but does not tell us what they are. However, some philosophers have made concrete suggestions about phenomena which they think science can never explain. An example is *consciousness*—the distinguishing feature of thinking, feeling creatures such as ourselves and other higher animals. Much research into the nature of consciousness has been and continues to be done, by neuroscientists, psychologists, and others. But a number of recent philosophers claim that whatever this research throws up, it will never fully explain the nature of consciousness. There is something intrinsically mysterious about the phenomenon of consciousness, they maintain, that no amount of scientific investigation can eliminate.

What are the grounds for this view? The basic argument is that conscious experiences are fundamentally unlike anything else in the world, in that they have a 'subjective aspect'. Consider for example the experience of watching a terrifying horror movie. This is an experience with a very distinctive 'feel' to it; in the current jargon, there is 'something that it is like' to have the experience. Neuroscientists may one day be able to give a detailed account of the complex goings-on in the brain which produce our feeling of terror. But will this explain why watching a horror movie feels the way it does, rather than feeling some other way? Some philosophers argue that it will not. On their view, the scientific study of the brain can at most tell us which brain processes are correlated with which conscious experiences. This is certainly interesting and valuable information. However it doesn't tell us *why* experiences with distinctive subjective 'feels' should result from the purely physical goings-on in the brain. Hence consciousness, or at least one important aspect of it, is scientifically inexplicable.

Though quite compelling, this argument is controversial and not endorsed by all philosophers, let alone all neuroscientists. Indeed a well-known 1991 book by the philosopher Daniel Dennett is defiantly entitled *Consciousness Explained*. Supporters of the view that consciousness is scientifically inexplicable are sometimes accused of having a lack of imagination. Even if it is true that brain science as currently practised cannot explain the subjective aspect of conscious experience, can we not imagine the emergence of a different type of brain science, with different explanatory techniques, that *does* explain why our experiences feel the way they do? There is a long tradition of philosophers trying to tell scientists what is and isn't possible, and later scientific developments have often proved the philosophers wrong. Only time will tell whether the same fate awaits those who argue that consciousness must always elude scientific explanation.

Explanation and reduction

The different scientific disciplines are designed for explaining different types of phenomena. To explain why rubber doesn't conduct electricity is a task for physics. To explain why turtles have such long lives is a task for biology. To explain why higher interest rates reduce inflation is a task for economics, and so on. In short, there is a division of labour between the different sciences: each specializes in explaining its own particular set of phenomena. This explains why the sciences are not usually in competition with one another—why biologists, for example, do not worry that physicists and economists might encroach on their turf.

Nonetheless, it is widely held that the different branches of science are not all on a par: some are more fundamental than others. Physics is usually regarded as the most fundamental science of all. Why? Because the objects studied by the other sciences are ultimately composed of physical particles. Consider living organisms, for example. Living organisms are made up of cells, which are themselves made up of water, nucleic acids, proteins, sugars, and lipids, all of which consist of molecules or long chains of molecules joined together. But molecules are made up of atoms, which are physical particles. So the objects biologists study are ultimately just very complex physical entities. The same applies to the other sciences, even the social sciences. Take economics for example. Economics studies the behaviour of firms and consumers in the market place, and the consequences of this behaviour. But consumers are human beings and firms are made up of human beings; and human beings are living organisms, hence physical entities.

Does this mean that, in principle, physics can subsume all the higher-level sciences? Since everything is made up of physical particles, surely if we had a complete physics, which allowed us to predict perfectly the behaviour of every physical particle in the universe, all the other sciences would become superfluous? Most philosophers resist this line of thought. After all, it seems crazy to suggest that physics might one day be able to explain the things that biology and economics explain. The prospect of deducing the laws of biology and economics straight from the laws of physics looks very remote. Whatever the physics of the future looks like, it is most unlikely to be capable of predicting economic downturns. Far from being reducible to physics, sciences such as biology and economics seem largely autonomous of it.

This leads to a philosophical puzzle. How can a science which studies entities which are ultimately physical *not* be reducible to physics? Granted that the higher-level sciences are in fact autonomous of physics, how is this possible? According to some philosophers, the answer lies in the fact that the objects studied by the higher-level sciences are *multiply realized* at the physical level. To illustrate the idea of multiple realization, imagine a collection of ashtrays. Each individual ashtray is obviously a physical entity, like everything else in the universe. But the physical composition of the ashtrays could be very different—some might be made of glass, others of aluminium, others of plastic, and so on. And they will probably differ in size, shape, and weight. There is virtually no limit on the range of different physical properties that an ashtray can have. So it is impossible to define the concept 'ashtray' in purely physical terms. We cannot find a true statement of the form 'x is an ashtray if and only if x is…' where the blank is filled by an expression taken from the language of physics. This means that ashtrays are multiply realized at the physical level.

Philosophers have often invoked multiple realization to explain why psychology cannot be reduced to physics or chemistry, but in principle the explanation works for any higher-level science. For example, consider the biological fact that nerve cells live longer than skin cells. Cells are physical entities, so one might think that this fact will one day be explained by physics. However, cells are almost certainly multiply realized at the microphysical level. Cells are ultimately made up of atoms, but the precise arrangement of atoms will be very different in different cells. So the concept 'cell' cannot be defined in terms drawn from fundamental physics. There is no true statement of the form 'x is a cell if and only if x is . . .' where the blank is filled by an expression taken from the language of microphysics. If this is correct, it means that fundamental physics will never be able to explain why nerve cells live longer than skin cells, or indeed any other facts about cells. The vocabulary of cell biology and the vocabulary of physics do not map onto each other in the required way. Thus we have an explanation of why it is that cell biology cannot be reduced to physics, despite the fact that cells are physical entities. Not all philosophers are happy with the doctrine of multiple realization, but it does promise a neat explanation of the autonomy of the higher-level sciences, both from physics and from each other.