PHILOSOPHICAL ISSUES IN SCIENCE SERIES
Edited by W. H. Newton-Smith

THE RATIONAL AND THE SOCIAL
James Robert Brown

THE NATURE OF DISEASE
Lawrie Reznek

THE PHILOSOPHICAL DEFENCE OF PSYCHIATRY
Lawrie Reznek

INFERENCE TO THE BEST EXPLANATION
Peter Lipton

TIME, SPACE AND PHILOSOPHY
Christopher Ray

MATHEMATICS AND THE IMAGE OF REASON
Mary Tiles

METAPHYSICS OF CONSCIOUSNESS
William Seager

INFERENCE TO THE BEST EXPLANATION

Peter Lipton

ROUTLEDGE
London and New York

1

INDUCTION

UNDERDETERMINATION

Inductive inference is a matter of weighing evidence and judging likelihood, not of proof. How do we go about making these non-demonstrative judgments, and why should we believe they are reliable? Both the question of description and the question of justification arise from underdetermination. To say that an outcome is underdetermined is to say that some information about initial conditions and rules or principles does not guarantee a unique solution. The information that Tom spent five dollars on apples and oranges and that apples are fifty cents a pound and oranges a dollar a pound underdetermines how much fruit Tom bought, given only the rules of deduction. Similarly, those rules and a finite number of points on a curve underdetermine the curve, since there are many curves that would pass through those points.

Underdetermination may also arise in our description of the way a person learns or makes inferences. A description of the evidence, along with a certain set of rules, not necessarily just those of deduction, may underdetermine what is learned or inferred. Insofar as we have described all the evidence and the person is not behaving erratically, this shows that there are hidden rules. We can then study the patterns of learning or inference to try to discover them. Noam Chomsky's argument from 'the poverty of the stimulus' is a good example of how underdetermination can be used to disclose the existence of additional rules (1965, ch. 1, sec. 8, esp. pp. 58–9). Children learn the language of their elders, an ability that enables them to understand an indefinite number of sentences on first acquaintance. The talk

young children hear, however, along with rules of deduction and any plausible general rules of induction, grossly underdetermine the language they learn. What they hear is limited and often includes many ungrammatical sentences, and the little they hear that is well formed is compatible with many possible languages other than the one they learn. Therefore, Chomsky argues, in addition to any general principles of deduction and induction, children must be born with strong linguistic rules or principles that further restrict the class of languages they will learn, so that the actual words they hear are now sufficient to determine a unique language. Moreover, since a child will learn whatever language he is brought up in, these principles cannot be peculiar to a particular human language; instead, they must specify something that is common to all of them. For Chomsky, determining the structure of these universal principles and the way they work is the central task of modern linguistics.

Thomas Kuhn provides another well-known example of using underdetermination as a tool to investigate cognitive principles. He begins from an argument about scientific research strikingly similar to Chomsky's argument about language acquisition (1970; 1977, esp. ch. 12). In most periods in the history of a developed scientific specialty, scientists are in broad agreement about which problems to work on, how to attack them, and what counts as solving them. But the explicit beliefs and rules scientists share, especially their theories, data, general rules of deduction and induction, and any explicit methodological rules, underdetermine these shared judgments. Many possible judgments are compatible with these beliefs and rules other than the ones the scientists make. So Kuhn argues that there must be additional field-specific principles that determine the actual judgments. Unlike Chomsky, Kuhn does not argue for principles that are either innate or in the form of rules, narrowly construed. Instead, scientists acquire through their education a stock of exemplars – concrete problem solutions in their specialty – and use them to guide their research. They pick new problems that look similar to an exemplar problem, they try techniques that are similar to those that worked in that exemplar, and they assess their success by reference to the standards of solution that the exemplars illustrate. Thus the exemplars set up a web of 'perceived similarity relations' that guide future research, and the shared judgments are explained by the shared exemplars. These similarities are not created or governed by rules, but they result in a pattern of research that roughly mimics one that is rule-governed. Just how exemplars do this work, and what happens when they stop working, provide the focus of Kuhn's account of science.

Chomsky and Kuhn are both arguing for unacknowledged principles of induction. In these cases, however, underdetermination is taken to be a symptom of the existence of highly specialized principles, whether of language acquisition or of scientific method in a particular field, since the underdetermination is claimed to remain even if we include general principles of induction among our rules. But the same pattern of argument applies to the general case. If an inference is inductive, then by definition it is underdetermined by the evidence and the rules of deduction. Insofar as our inductive practices are methodical, we must use additional rules or principles of inference, and we may study the patterns of our inferences in an attempt to discover them.

JUSTIFICATION

The two central questions about our general principles of induction concern description and justification. What principles do we use, and how can they be shown to be good principles? The question of description seems at first to take priority. How can we even attempt to justify our principles until we know what they are? Historically, however, the justification question came first. One reason for this is that this question gets its grip from skeptical arguments that seem to apply to any principles that could account for the way we fill the gap between evidence and inference.

The problem of justification is to show that our inferential methods are good methods. The natural way to understand this is in terms of truth. We want our methods of inference to be 'truth-tropic', to take us towards the truth. For deduction, a good argument is one that is valid, a perfect truth conduit, where it is impossible for there to be true premises but a false conclusion. The problem of justification here would be to show that arguments we judge valid are in fact so. For induction, perfect reliability is out of the question. By definition, even a good inductive argument is one where it is possible for there to be true premises but a false conclusion. Moreover, it is clear that the

reasonable inductive inferences we make are not 100 per cent reliable even in this world, since they sometimes sadly take us from truth to falsehood. Nevertheless, it remains natural to construe the task of justification as that of showing truth-tropism. We would like to show that those inductive inferences we judge worth making are ones that usually take us from true premises to true conclusions.

A skeptical argument that makes the problem of justification pressing has two components, underdetermination and circularity. The first is an argument that the inferences in question are underdetermined, given only our premises and the rules of deduction; that the premises and those rules are compatible not just with the inferences we make, but also with other, incompatible inferences. This shows that the inferences we make really are inductive and, by showing that there are possible worlds where the principles we use take us from true premises to false conclusions, it also shows that there are worlds where our principles would fail us. Revealing this underdetermination, however, does not yet generate a skeptical argument, since we might have good reason to believe that the actual world is one where our principles are reliable. So the skeptical argument requires a second component, an argument for circularity, which attempts to show that we cannot rule out the possibility of unreliability that underdetermination raises without employing the very principles that are under investigation, and so begging the question.

Although it is not traditionally seen as raising the problem of induction, Descartes' 'First Meditation' is a classic illustration of this technique. Descartes' goal is to cast doubt on the 'testimony of the senses', which leads us to infer that there is, say, a mountain in the distance because that is what it looks like. He begins by arguing that we ought not to trust the senses completely, since we know that they do sometimes mislead us, 'when it is a question of very small and distant things'. This argument relies on the underdetermination component, since it capitalizes on the fact that the way things appear does not entail the way they are, but it does not yet have the circle component. We can corroborate our inferences about small and distant things without circularity by taking a closer look (cf. Williams, 1978, pp. 51–2). But Descartes immediately moves on from the small and the distant to the large and near. No matter how clearly we seem to see something, it may only be a dream, or a misleading experience induced by an evil demon. These arguments describe possible situations where even the most obvious testimony of the senses is misleading. Moreover, unlike the worry about small and distant things, these arguments also have a circle component. There is apparently no way to test whether a demon is misleading us with a particular experience, since any test would itself rely on experiences that he might have induced. The senses may be liars and give us false testimony, and we should not find any comfort if they also report that they are telling us the truth.

The demon argument begins with the underdetermination of observational belief by observational experience, construes the missing principle of inference on the model of inference from testimony, and then suggests that the reliability of this principle could only be shown by assuming it. Perhaps one of the reasons Descartes' arguments are not traditionally seen as raising the problem of justifying induction is that his response to his own skepticism is to reject the underdetermination upon which it rests. Descartes argues that, since inferences from the senses must be inductive and so raise a skeptical problem, our knowledge must instead have a different sort of foundation for which the problem of underdetermination does not arise. The *cogito* and the principles of clearness and distinctness that it exemplifies are supposed to provide the non-inductive alternative. In other words, the moral he draws from underdetermination and circularity is not that our principles of induction require some different sort of defense or must be accepted without justification, but that we must use different premises and principles, for which the skeptical problem does not arise. For a skeptical argument about induction that does not lead to the rejection of induction, we must turn to its traditional home, in the arguments of David Hume.

Hume also begins with underdetermination, in this case that our observations do not entail our predictions (1777, sec. IV). He then suggests that the governing principle of all our inductive inferences is that nature is uniform, that the unobserved (but observable) world is much like what we have observed. The question of justification is then the question of showing that nature is indeed uniform. This cannot be deduced from what we have observed, since the claim of uniformity itself incorporates a massive prediction. But the only other way to argue for

uniformity is to use an inductive argument, which would rely on the principle of uniformity, leaving the question begged. According to Hume, we are addicted to the practice of induction, but it is a practice that cannot be justified.

To illustrate the problem, suppose our fundamental principle of inductive inference is 'More of the Same'. We believe that strong inductive arguments are those whose conclusions predict the continuation of a pattern described in the premises. Applying this principle of conservative induction, we would infer that the sun will rise tomorrow, since it has always risen in the past; and we would judge worthless the argument that the sun will not rise tomorrow since it has always risen in the past. It is, however, easy to come up with a factitious principle to underwrite the latter argument. According to the principle of revolutionary induction, 'It's Time for a Change', and this sanctions the dark inference. Hume's argument is that we have no way to show that conservative induction, the principle he claims we actually use for our inferences, will do any better than intuitively wild principles like the principle of revolutionary induction. Of course conservative induction has had the more impressive track record. Most of the inferences from true premises that it has sanctioned have also had true conclusions. Revolutionary induction, by contrast, has been conspicuous in failure, or would have been, had anyone relied on it. The question of justification, however, does not ask which method of inference has been successful; it asks which one will be successful.

Still, the track record of conservative induction appears to be a reason to trust it. That record is imperfect (we are not aspiring to deduction), but very impressive, particularly as compared with revolutionary induction and its ilk. In short, induction will work because it has worked. This seems the only justification our inductive ways could ever have or require. Hume's disturbing observation was that this justification appears circular, no better than trying to convince someone that you are honest by saying that you are. Much as Descartes argued that we should not be moved if the senses give testimony on their own behalf, so Hume argued that we cannot appeal to the history of induction to certify induction. The trouble is that the argument that conservative inductions will work because they have worked is itself an induction. The past success is not supposed to prove future success, only make it very likely. But then we must decide which standards to use to evaluate this argument. It has the form 'More of the Same', so conservatives will give it high marks, but since its conclusion is just to underwrite conservatism, this begs the question. If we apply the revolutionary principle, it counts as a very weak argument. Worse still, by revolutionary standards, conservative induction is likely to fail precisely because it has succeeded in the past, and the past failures of revolutionary induction augur well for its future success (cf. Skyrms, 1986, ch. 2). The justification of revolutionary induction seems no worse than the justification of conservative induction, which is to say that the justification of conservative induction looks very bad indeed.

The problem of justifying induction does not show that there are other inductive standards better than our own. Instead it argues for a deep symmetry: many sets of standards, most of them wildly different from our own and incompatible with each other, are yet completely on a par from a justificatory point of view. This is why the problem of justification can be posed before we have solved the problem of description. Whatever inductive principles we use, the fact that they are inductive seems enough for the skeptic to show that they defy justification. We fill the gap of underdetermination between observation and prediction in one way, but it could be filled in many other ways that would have led to entirely different predictions. We have no way of showing that our way is any better than any of these others. It is not merely that the revolutionaries will not be convinced by the justificatory arguments of the conservatives: the conservatives should not accept their own defense either, since among their standards is one which says that a circular argument is a bad argument, even if it is in one's own aid. Even if I am honest, I ought to admit that the fact that I say so ought not carry any weight. We have a psychological compulsion to favor our own inductive principles but, if Hume is right, we should see that we cannot even provide a cogent rationalization of our behavior.

It seems to me that we do not yet have a satisfying solution to Hume's challenge and that the prospects for one are bleak, but there are other problems of justification that are more tractable. The peculiar difficulty of meeting Hume's skeptical argument against induction is that he casts doubt on our inductive principles as a whole, and so any recourse to induction to justify induction seems hopeless. But one can also ask for the

justification of particular inductive principles and, as Descartes' example of small and distant things suggests, this leaves open the possibility of appeal to other principles without begging the question. For example, among our principles of inference is one that makes us more likely to infer a theory if it is supported by a variety of evidence than if it is supported by a similar amount of homogeneous data. This is the sort of principle that might be justified in terms of a more basic inductive principle, say that we have better reason to infer a theory when all the reasonable competitors have been refuted, or that a theory is only worth inferring when each of its major components has been separately tested. Another, more controversial, example of a special principle that might be justified without circularity is that, all else being equal, a theory deserves more credit from its successful predictions than it does from data that the theory was constructed to fit. This appears to be an inductive preference most of us have, but the case is controversial because it is not at all obvious that it is rational. On the one hand, many people feel that only a prediction can be a real test, since a theory cannot possibly be refuted by data it is built to accommodate; on the other, that logical relations between theory and data upon which inductive support exclusively depends cannot be affected by the merely historical fact that the data were available before or only after the theory was proposed. In any event, this is an issue of inductive principle that is susceptible to non-circular evaluation, as we will see in chapter eight. Finally, though a really satisfying solution to Hume's problem would have to be an argument for the reliability of our principles that had force against the inductive skeptic, there may be arguments for reliability that do not meet this condition yet still have probative value for those of us who already accept some forms of induction. We will consider some candidates in chapter nine.

DESCRIPTION

We can now see why the problem of justification, the problem of showing that our inductive principles are reliable, did not have to wait for a detailed description of those principles. The problem of justifying our principles gets its bite from skeptical arguments, and these appear to depend only on the fact that these principles are principles of induction, not on the particular form they take.

The crucial argument is that the only way to justify our principles would be to reason with the very same principles, which is illegitimate; an argument that seems to work whatever the details of our inferences. The irrelevance of the details comes out in the symmetry of Hume's argument: just as the future success of conservative induction gains no plausibility from its past success, so the future success of revolutionary induction gains nothing from its past failures. As the practice varies, so does the justificatory argument, preserving the pernicious circularity. Thus the question of justification has had a life of its own: it has not waited for a detailed description of the practice whose warrant it calls into doubt.

By the same token, the question of description has fortunately not waited for an answer to the skeptical arguments. Even if our inferences were unjustifiable, one still might be interested in saying how they work. The problem of description is not to show that our inferential practices are reliable; it is simply to describe them as they stand. One might have thought that this would be a relatively trivial problem. First of all, there are no powerful reasons for thinking that the problem is impossible, as there are for the problem of justification. There is no great skeptical argument against the possibility of description. It is true that any account of our principles will require inductive support, since we must see whether it jibes with our observed inductive practice. This, however, raises no general problem of circularity now that justification is not the issue. Using induction to investigate induction is no more a problem here than using observation to study the structure and function of the eye. Second, it is not just that a solution to the problem of describing our inductive principles should be possible, but that it should be easy. After all, they are our principles, and we use them constantly. It thus comes as something of a shock to discover how extraordinarily difficult the problem of description has turned out to be. It is not merely that ordinary reasoners are unable to describe what they are doing: at least one hundred and fifty years of focused effort by epistemologists and philosophers of science has yielded little better. Again, it is not merely that we have yet to capture all the details, but that the most popular accounts of the gross structure of induction are wildly at variance with our actual practice.

Why is description so hard? One reason is a quite general gap between what we can do and what we can describe. You may

know how to do something without knowing how you do it; indeed, this is the usual situation. It is one thing to know how to tie one's shoes or to ride a bike; it is quite another thing to be able to give a principled description of what it is that one knows. Chomsky's work on principles of language acquisition and Kuhn's work on scientific method are good cognitive examples. Their investigations would not be so important and controversial if the ordinary speaker of a language knew how she distinguished grammatical from ungrammatical sentences or the normal scientist knew how he made his methodological judgments. The speaker and the scientist employ various principles, but they are not conscious of them. The situation is similar in the case of inductive inference generally. Although we may partially articulate some of our inferences if, for example, we are called upon to defend them, we are not conscious of the fundamental principles of inductive inference we constantly use.

Since our principles of induction are neither available to introspection, nor observable in any other way, the evidence for their structure must be highly indirect. The project of description is one of black box inference, where we try to reconstruct the underlying mechanism on the basis of the superficial patterns of evidence and inference we observe in ourselves. This is no trivial problem. Part of the difficulty is simply the fact of underdetermination. As the examples of Chomsky and Kuhn show, underdetermination can be a symptom of missing principles and a clue to their nature, but it is one that does not itself determine a unique answer. In other words, where the evidence and the rules of deduction underdetermine inference, that information also underdetermines the missing principles. There will always be many different possible mechanisms that would produce the same patterns, so how can one decide which one is actually operating? In practice, however, we usually have the opposite problem: we can not come up with even one description that would yield the patterns we observe. The situation is the same in scientific theorizing generally. There is always more than one account of the unobserved and often unobservable world that would account for what we observe, but scientists' actual difficulty is often to come up with even one theory that fits the observed facts. On reflection, then, it should not surprise us that the problem of description has turned out to be so difficult. Why should we suppose that the project of describing our inductive principles is going to be easier than it would be, say, to give a detailed account of the working of a computer on the basis of the correlations between keys pressed and images on the screen?

Now that we are prepared for the worst, we may turn to some of the popular attempts at description. In my discussion of the problem of justification, I suggested that, following Hume's idea of induction as habit formation, we describe our pattern of inference as 'More of the Same'. This is pleasingly simple, but the conservative principle is at best a caricature of our actual practice. We sometimes do not infer that things will remain the same and we sometimes infer that things are going to change. When my mechanic tells me that my brakes are about to fail, I do not suppose that he is therefore a revolutionary inductivist. Again, we often make inductive inferences from something we observe to something invisible, such as from people's behavior to their beliefs or from the scientific evidence to unobservable entities and processes, and this does not fit into the conservative mold. 'More of the Same' might enable me to predict what you will do on the basis of what you have done (if you are a creature of habit), but it will not tell me what you are or will be thinking.

Faced with the difficulty of providing a general description, a reasonable strategy is to begin by trying to describe one part of our inductive practice. This is a risky procedure, since the part one picks may not really be describable in isolation, but there are sometimes reasons to believe that a particular part is independent enough to permit a useful separation. Chomsky must believe this about our principles of linguistic inference. Similarly, one might plausibly hold that, while simple habit formation cannot be the whole of our inductive practice, it is a core mechanism that can be treated in isolation. Thus one might try to salvage the intuition behind the conservative principle by giving a more precise account of the cases where we are willing to project a pattern into the future, leaving to one side the apparently more difficult problems of accounting for predictions of change and inferences to the unobservable. What we may call the instantial model of inductive confirmation may be seen in this spirit. According to it, a hypothesis of the form 'All A's are B' is supported by its positive instances, by observed A's that are also B (cf. Hempel, 1965, ch. 1). This is not, strictly speaking, an account of inductive *inference*, since it does not say either how we come up with the hypothesis in the first place or how many supporting instances are required

before we actually infer it, but this switching of the problem from inference to support may also be taken as a strategic simplification. In any event, the underlying idea is that, if enough positive instances and no refuting instances (A's that are not B) are observed, we will infer the hypothesis, from which we may then deduce the prediction that the next A we observe will be B.

This model could only be a very partial description of our inductive principles but, within its restricted range, it strikes many people initially as a truism, and one that captures Hume's point about our propensity to extend observed patterns. Observed positive instances are not necessary for inductive support, as inferences to the unobserved and to change show, but they might seem at least sufficient. The instantial model, however, has been shown to be wildly over-permissive. Some hypotheses are supported by their positive instances, but many are not. Observing only black ravens may lead one to believe that all ravens are black, but observing only bearded philosophers would probably not lead one to infer that all philosophers are bearded. Nelson Goodman has generalized this problem, by showing how the instantial model sanctions any prediction at all if there is no restriction on the hypotheses to which it can be applied (Goodman, 1983, ch. 3). His technique is to construct hypotheses with factitious predicates. Black ravens provide no reason to believe that the next swan we see will be white, but they do provide positive instances of the artificial hypothesis that 'All raveswans are blight', where something is a raveswan just in case it is either observed before today and a raven, or not so observed and a swan, and where something is blight just in case it is either observed before today and black, or not so observed and white. But the hypothesis that all raveswans are blight entails that the next observed raveswan will be blight which, given the definitions, is just to say that the next swan will be white.

The other famous difficulty facing the instantial model arises for hypotheses that do seem to be supported by their instances. Black ravens support the hypothesis that all ravens are black. This hypothesis is logically equivalent to the contrapositive hypothesis that all non-black things are non-ravens: there is no possible situation where one hypothesis would be true but the other false. According to the instantial model, the contrapositive hypothesis is supported by non-black non-ravens, such as green leaves. The rub comes with the observation that whatever supports a hypothesis also supports anything logically equivalent to it. This is very plausible, since support provides a reason to believe true, and we know that if a hypothesis is true, then so must be anything logically equivalent to it. But then the instantial model once again makes inductive support far too easy, counting green leaves as evidence that all ravens are black (Hempel, 1965, ch. 1). We will discuss this paradox of the ravens in chapter six.

Another account of inductive support is the hypothetico-deductive model (Hempel, 1966, chs 2, 3). On this view, a hypothesis or theory is supported when it, along with various other statements, deductively entails a datum. Thus a theory is supported by its successful predictions. This account has a number of attractions. First, although it leaves to one side the important question of the source of hypotheses, it has much wider scope than the instantial model, since it allows for the support of hypotheses that appeal to unobservable entities and processes. The big bang theory of the origin of the universe obviously cannot be directly supported but, along with other statements, it entails that we ought to find ourselves today traveling through a uniform background radiation, like the ripples left by a rock falling into a pond. The fact that we do now observe this radiation (or effects of it) provides some reason to believe the big bang theory. Thus, even if a hypothesis cannot be supported by its instances, because its instances are not observable, it can be supported by its observable logical consequences. Second, the model enables us to co-opt our accounts of deduction for an account of induction, an attractive possibility since our understanding of deductive principles is so much better than our understanding of inductive principles. Lastly, the hypothetico-deductive model seems genuinely to reflect scientific practice, which is perhaps why it has become the scientists' philosophy of science.

In spite of all its attractions, our criticism of the hypothetico-deductive model here can be brief, since it inherits all the over-permissiveness of the instantial model. Any case of support by positive instances will also be a case of support by consequences. The hypothesis that all A's are B, along with the premise that an individual is A, entails that it will also be B, so the thing observed to be B supports the hypothesis, according to the hypothetico-deductive model. That is, any case of instantial support is also a

case of hypothetico-deductive support, so the model has to face the problem of insupportable hypotheses and the raven paradox. Moreover, the hypothetico-deductive model is similarly over-permissive in the case of vertical inferences to hypotheses about unobservables, a problem that the instantial model avoided by ignoring such inferences altogether. The difficulty is structurally similar to Goodman's problem of factitious predicates. Consider the conjunction of the hypotheses that all ravens are black and that all swans are white. This conjunction, along with premises concerning the identity of various ravens, entails that they will be black. According to the model, the conjunction is supported by black ravens, and it entails its own conjunct about swans. The model thus appears to sanction the inference from black ravens to white swans (cf. Goodman, 1983, pp. 67–8). Similarly, the hypothesis that all swans are white taken alone entails the inclusive disjunction that either all swans are white or there is a black raven, a disjunction we could establish by seeing a black raven, thus giving illicit hypothetico-deductive support to the swan hypothesis. These maneuvers are obviously artificial, but nobody has managed to show how the model can be modified to avoid them without also eliminating most genuine cases of inductive support (cf. Glymour, 1980, ch. 2). Finally, in addition to being too permissive, finding support where none exists, the model is also too strict, since data may support a hypothesis which does not, along with reasonable auxiliary premises, entail them. We will investigate this problem in chapter five and return to the problem of over-permissiveness in chapter six.

We have now canvassed three attempts to tackle the descriptive problem: 'More of the Same', the instantial model, and the hypothetico-deductive model. The first two could, at best, account for a special class of particularly simple inferences, and all three are massively over-permissive, finding inductive support where there is none to be had. They do not give enough structure to the black box of our inductive principles to determine the inferences we actually make. This is not to say that these accounts describe mechanisms that would yield too many inferences: they would probably yield too few. A 'hypothetico-deductive box', for example, would probably have little or no inferential output, given the plausible additional principle that we will not make inferences we know to be contradictory. For every hypothesis that we would be inclined to infer on the basis of the deductive support it enjoys, there will be an incompatible hypothesis that is similarly supported, and the result is no inference at all, so long as both hypotheses are considered and their incompatibility recognized.

A fourth account of induction, the last I will consider in this section, focuses on causal inference. It is a striking fact about our inductive practice, both lay and scientific, that so many of our inferences depend on inferring from effects to their probable causes. This is something that Hume himself emphasized (Hume, 1777, sec. IV). The cases of direct causal inference are legion, such as the doctor's inference from symptom to disease, the detective's inference from evidence to perpetrator, the mechanic's inference from the engine noises to what is broken, and many scientific inferences from data to theoretical explanation. Moreover, it is striking that we often make a causal inference even when our main interest is in prediction. Indeed, the detour through causal theory on the route from data to prediction seems to be at the center of many of the dramatic successes of scientific prediction. All this suggests that we might do well to consider an account of the way causal inference works as a central component of a description of our inductive practice.

The best known account of causal inference is John Stuart Mill's discussion of the 'methods of experimental inquiry' (Mill, 1904, book III, ch. VIII). The two central methods are the Method of Agreement and especially the Method of Difference. According to the Method of Agreement, in idealized form, when we find that there is only one antecedent that is shared by all the observed instances of an effect, we infer that it is a cause (book III, ch. VIII, sec. 1; references to Mill hereafter as, e.g., 'III.VIII.1'). Thus we come to believe that hangovers are caused by heavy drinking. According to the Method of Difference, when we find that there is only one prior difference between a situation where the effect occurs and an otherwise similar situation where it does not, we infer that the antecedent that is only present in the case of the effect is a cause (III.VIII.2). If we add sodium to a blue flame, and the flame turns yellow, we infer that the presence of sodium is a cause of the new color, since that is the only difference between the flame before and after the sodium was added. If we once successfully follow a recipe for baking bread, but fail another time because we have left out the yeast and the bread does not rise, we would infer that the yeast is a cause of the rising in the

first case. Both methods work by a combination of retention and variation. When we apply the Method of Agreement, we hold the effect constant and try to vary the background as much as we can, and see what stays the same; when we apply the Method of Difference, we vary the effect, and try to hold as much of the background constant as we can, and see what changes.

Mill's methods have a number of attractive features. Many of our inferences are causal inferences, and Mill's methods give a natural account of these. In science, for example, the controlled experiment is a particularly common and self-conscious application of the Method of Difference. The Millian structure of causal inference is often particularly clear in cases of inferential dispute. When you dispute my claim that C is the cause of E, you will often make your case by pointing out that the conditions for Mill's methods are not met; that is, by pointing out C is not the only antecedent common to all cases of E, or that the presence of C is not the only salient difference between a case where E occurs and a similar case where it does not. Mill's methods may also avoid some of the over-permissiveness of other accounts, because of the strong constraints that the requirements of varied or shared backgrounds place on their application. These requirements suggest how our background beliefs influence our inferences, something a good account of inference must do. The methods also help to bring out the roles in inference of competing hypotheses and negative evidence, as we will see in chapter five, and the role of background knowledge, as we will see in chapter seven.

Of course, Mill's methods have their share of liabilities, of which I will mention just two. First, they do not themselves apply to unobservable causes or to any causal inferences where the cause's existence, and not just its causal status, is inferred. Second, if the methods are to apply at all, the requirement that there be only a single agreement or difference in antecedents must be seen as an idealization, since this condition is never met in real life. We need principles for selecting from among multiple agreements or similarities those that are likely to be causes, but these are principles Mill does not himself supply. As we will see in chapters five through seven, however, Mill's method can be modified and expanded in a way that may avoid these and other liabilities it faces in its simple form.

This chapter has set part of the stage for an investigation of our inductive practices. I have suggested that many of the problems those practices raise can be set out in a natural way in terms of the underdetermination that is characteristic of inductive inference. The underdetermination of our inferences by our evidence provides the skeptic with his lever, and so poses the problem of justification. It also elucidates the structure of the descriptive problem, and the black box inferences it will take to solve it. I have canvassed several solutions to the problem of description, partly to give a sense of some of our options and partly to suggest just how difficult the problem is. But at least one solution to the descriptive problem was conspicuous by its absence, the solution that gives this book its title and which will be at the center of attention from chapter four onwards. According to Inference to the Best Explanation, we infer what would, if true, be the best explanation of our evidence. On this view, explanatory considerations are our guide to inference. So to develop and assess this view, we need first to look at another sector of our cognitive economy, our explanatory practices. This is the subject of the next two chapters.

2

EXPLANATION

UNDERSTANDING EXPLANATION

Once we have made an inference, what do we do with it? Our inferred beliefs are guides to action that help us to get what we want and avoid trouble. Less practically, we also sometimes infer simply because we want to learn more about the way the world is. Often, however, we are not satisfied to discover that something is the case: we want to know *why*. Thus our inferences may be used to provide explanations, and they may themselves be explained. The central question about our explanatory practices can be construed in several ways. We may ask what principles we use to distinguish between a good explanation, a bad explanation, and no explanation at all. Or we may ask what relation is required between two things to count one to be an explanation of the other. We can also formulate the question in terms of the relationship between knowledge and understanding. Typically, someone who asks why something is the case already knows that it is the case. The person who asks why the sky is blue knows that it is blue, but does not yet understand why. The question about explanation can then be put this way: What has to be added to knowledge to get understanding?

As in the case of inference, explanation raises problems both of justification and of description. The problem of justification can be understood in various ways. It may be seen as the problem of showing whether things we take to be good explanations really are, whether they really provide understanding. The issue here, to distinguish it from the case of inference, is not whether there is any reason to believe that our putative explanations are themselves true, but whether, granting that they are true, they really