# 4

# INFERENCE TO THE BEST EXPLANATION

## SPELLING OUT THE SLOGAN

Our initial survey of the problems of induction and explanation is now complete. We have considered some of the forms these problems take, some of the reasons they are so difficult to solve, and some of the weaknesses in various attempts to solve them. In the last chapter, I also attempted something more constructive, by giving what I hope is an improved version of the causal model of explanation. Up to now, however, we have treated inference and explanation virtually in mutual isolation, a separation that reflects most of the literature on these subjects. Although the discussion of inference in chapter one construed the task of describing our practices as itself an explanatory inquiry, the attempt to specify the black box mechanism that takes us from evidence to inference and so explains why we make the inferences we do, none of the models of inference we considered explicitly invoked explanatory relations between evidence and conclusion. Similarly, in the discussion of explanation in chapters two and three, inferential considerations played a role in only one of the models, the reason model. That model uses an inferential notion to account for explanation, by claiming that we explain a phenomenon by giving some reason to believe that the phenomenon occurs, and it was found to be unacceptable for inferential reasons, since it does not allow for self-evidencing explanations, such as the explanation of the tracks in the snow or of the red shift of the star, where the phenomenon that is explained provides an essential part of the reason for believing that the explanation is correct.

In this chapter, the relationship between the practices of explanation and of inference will take center stage, where it will

remain for the rest of this book. Let us begin with a very simple view of that relationship. First we make our inferences; then, when we want to explain a phenomenon, we draw upon our pool of beliefs for an explanation, a pool filled primarily by those prior inferences. This, however, must be too simple, since our pool may not contain the explanation we seek. So a slightly less simple view is that, if we don't find an explanation in our pool, we search for a warranted inference that will explain, a process that may also require further observation. Explanatory considerations thus have some bearing on inference, since they may focus our inquiry but, on this view, inference still comes before explanation. After all, the most basic requirement of an explanation is that the explanatory information be correct, so how can we be in a position to use that information for an explanation unless we first know that it is correct?

This picture, however, seriously underestimates the role of explanatory considerations in inference. Those considerations tell us not only what to look for, but also whether we have found it. Take the cases of self-evidencing explanations. The tracks in the snow are the evidence for what explains them, that a person passed by on snowshoes; the red shift of the star is an essential part of the reason we believe the explanation, that it has a certain velocity of recession. In these cases, it is not simply that the phenomena to be explained provide reasons for inferring the explanations: we infer the explanations precisely because they would, if true, explain the phenomena. Of course, there is always more than one possible explanation for any phenomenon – the tracks might instead have been caused by a trained monkey on snowshoes, or by the elaborate etchings of an environmental artist – so we cannot infer something simply because it is a possible explanation. It must somehow be the best of competing explanations. These sorts of explanatory inferences are extremely common. The sleuth infers that the butler did it, since this is the best explanation of the evidence before him. The doctor infers that his patient has measles, since this is the best explanation of the symptoms. The astronomer infers the existence and motion of Neptune, since that is the best explanation of the observed perturbations of Uranus. Chomsky infers that our language faculty has a particular structure because this provides the best explanation of the way we learn to speak. Kuhn infers that normal science is governed by exemplars, since they provide the best

explanation for the observed dynamics of research. This suggests a new model of induction, one that binds explanation and inference in an intimate and exciting way. According to Inference to the Best Explanation, our inferential practices are governed by explanatory considerations. Given our data and our background beliefs, we infer what would, if true, provide the best of the competing explanations we can generate of those data (so long as the best is good enough for us to make any inference at all).

Inference to the Best Explanation has become extremely popular in philosophical circles, discussed by many and endorsed without discussion by many more (For discussions, see, e.g., Pierce, 1931, 5.180-5.212, esp. 5.189; Harman, 1965; Brody, 1970; Hanson, 1972, ch. 4; Thagard, 1978; Cartwright, 1983, essay 5). Yet it still remains much more of a slogan than an articulated account of induction. In the balance of this section, I will make some first steps towards improving this situation. In the next section, we will consider the initial attractions of the view, as well as some apparent liabilities. The balance of this book is devoted to the questions of whether Inference to the Best Explanation really will provide an illuminating model of our inductive practices and whether it is an improvement over the other accounts we have considered.

The obvious way to flesh out Inference to the Best Explanation would be to insert one of the standard models of explanation. This, however, yields disappointing results, because of the backward state of those models. For example, we wouldn't get very far if we inserted the deductive-nomological model, since this would just collapse Inference to the Best Explanation into a version of the hypothetico-deductive model of confirmation. Indeed one suitable acid test for Inference to the Best Explanation is that it mark an improvement over the hypothetico-deductive model. As we saw in chapter one, the deductive-nomological model of explanation has many unattractive features; it also provides almost no resources for saying when one explanation is better than another. We will do better with the causal model of contrastive explanation I developed in the last chapter, as we will see in chapters five and six, but for now we are better off not burdening Inference to the Best Explanation with the details of any specific model of explanation, trying instead to stick to the actual explanatory relation itself, whatever its correct description turns out to be. Let us begin to flesh out the account by developing two

signal distinctions that do not depend on the details of explanation: the distinction between actual and potential explanations, and the distinction between the explanation best supported by the evidence, and the explanation that would provide the most understanding or, in short, between the likeliest and the loveliest explanation.

Our discussion of inference, explanation, and the connection between the two is being conducted under the assumption of inferential and explanatory realism, an assumption we will not call into question until the final chapter of this book. I am assuming that a goal of inference is truth, that our actual inferential practices are truth-tropic, i.e. that they generally take us towards this goal, and that for something to be an actual explanation, it must be (at least approximately) true. But Inference to the Best Explanation cannot then be understood as inference to the best of the *actual* explanations. Such a model would make us too good at inference, since it would make all our inferences true. Our inductive practice is fallible: we sometimes reasonably infer falsehoods. This model would also fail to account for the role of competing explanations in inference. These competitors are typically incompatible and so cannot all be true, so we cannot represent them as competing actual explanations. The final and most important reason why Inference to the Best Actual Explanation could not describe our inductive practices is that it would not characterize the process of inference in a way we could follow, since we can only tell whether something is an actual explanation *after* we have settled the inferential question. It does not give us what we want, which is an account of the way explanatory considerations can serve as a guide to the truth. Telling someone to infer actual explanations is like a dessert recipe that says start with a soufflé. We are trying to describe the way we go from evidence to inference, but Inference to the Best Actual Explanation would require us already to have arrived in order to get there. In short, the model would not be epistemically effective.

The obvious solution, then, is to distinguish actual from *potential* explanation, and to construe Inference to the Best Explanation as Inference to the Best Potential Explanation. We have to produce a pool of potential explanations, from which we infer the best one. Although our discussion of explanation in the last two chapters considered only actual explanations, the distinction

between actual and potential explanations is familiar in the literature on explanation. The standard version of the deductive-nomological model gives an account of potential explanation: there is no requirement that the explanation be true, only that it include a general hypothesis and entail the phenomenon. If we then add a truth requirement, we get an account of actual explanation (Hempel 1965, p. 338). By shaving the truth requirement off explanation, we seem to get a notion suitable for Inference to the Best Explanation: one that allows for the distinction between warranted and successful inferences, permits the competition between explanations to take place among incompatible hypotheses, and gives an account that is epistemically effective. According to Inference to the Best Explanation, then, we do not infer the best actual explanation; rather we infer that the best of the available potential explanations is an actual explanation.

The intuitive idea of a potential explanation is of something that satisfies all the conditions of an actual explanation, except possibly that of truth (Hempel, 1965, p. 338). This characterization may, however, be somewhat misleading, since it seems to entail the false generalization that all true potential explanations are actual explanations. The generalization does not hold for explanations that fit the deductive-nomological model, since a lawlike statement could hold in one possible world as a law, but in another as a mere coincidence. Even more clearly, it does not hold in the context of a causal model. A potential cause may exist yet not be an actual cause, say because some other cause pre-empted it. Of course one could construct a technical notion of potential explanation that satisfied the equality between true potential explanation and actual explanation, but this would not be a suitable notion for Inference to the Best Explanation. As the literature on Gettier cases shows, we often infer potential causes that exist but are not actual causes (Gilbert Harman's two-candle case is a good example of this; see Harman, 1973, pp. 22–3).

So we may need to do more work to characterize the notion of potential explanation that is suitable for Inference to the Best Explanation. One issue is how large we should make the pool. We might say that a potential explanation is any account that is logically compatible with all our observations (or almost all of them) and that is a possible explanation of the relevant phenomena. In other words, the potential explanations of some

phenomena are those that do explain them in a possible world where our observations hold. This pool is very large, including all sorts of crazy explanations nobody would seriously consider. On the other hand, we might define the pool more narrowly, so that the potential explanations are only the 'live options': the serious candidates for an actual explanation. The advantage of the second characterization is that it seems to offer a better account of our actual procedure. When we decide which explanation to infer, we often start from a group of plausible candidates, and then consider which of these is the best, rather than selecting directly from the vast pool of possible explanations. In later chapters, we will see further reasons for preferring this restricted notion of potential explanation. But it is important to notice that the live options version of potential explanation already assumes an epistemic filter that limits the pool of potential explanations to plausible candidates. This version of Inference to the Best Explanation thus includes two filters, one that selects the plausible candidates, and a second that selects from among them. This view has considerable verisimilitude, but a strong version of Inference to the Best Explanation will not take the first filter as an unanalyzed mechanism, since epistemic filters are precisely the mechanisms that Inference to the Best Explanation is supposed to illuminate.

Let us turn now to the second distinction. It is important to distinguish two senses in which something may be the best of competing potential explanations. We may characterize it as the explanation that is most warranted: the 'likeliest' explanation. On the other hand, we may characterize the best explanation as the one which would, if correct, be the most explanatory or provide the most understanding: the 'loveliest' explanation. The criteria of likeliness and loveliness may well pick out the same explanation in a particular competition, but they are clearly different sorts of standard. Likeliness speaks of truth; loveliness of potential understanding. Moreover, the criteria do sometimes pick out different explanations. Sometimes the likeliest explanation is not very enlightening. It is extremely likely that smoking opium puts people to sleep because of its dormative powers (though not quite certain: it might be the oxygen that the smoker inhales with the opium, or even the depressing atmosphere of the opium den), but this is the very model of an unlovely explanation. An explanation can also be lovely without

being likely. Perhaps some conspiracy theories provide examples of this. By showing that many apparently unrelated events flow from a single source and many apparent coincidences are really related, such a theory may have considerable explanatory power. If only it were true, it would provide a very good explanation. That is, it is lovely. At the same time, such an explanation may be very unlikely, accepted only by those whose ability to weigh evidence has been tilted by paranoia.

One of the reasons that likeliness and loveliness sometimes diverge is that likeliness is relative to the total available evidence, while loveliness is not, or at least not in the same way. We may have an explanation that is both lovely and likely given certain evidence, unlikely given additional evidence, yet still a lovely explanation of the original evidence. Newtonian mechanics is one of the loveliest explanations in science and, at one time, it was also very likely. More recently, with the advent of special relativity and the new data that support it, Newtonian mechanics has become less likely, but it remains as lovely an explanation of the old data as it ever was. Another reason for the divergence is that the two criteria are differently affected by additional competition. A new competitor may decrease the likeliness of an old hypothesis, but it will usually not change its loveliness. Even without the evidence that favored special relativity, the production of the theory probably made Newtonian mechanics less likely but probably not less lovely.

This gives us two more versions of Inference to the Best Explanation to consider: Inference to the Likeliest Potential Explanation and Inference to the Loveliest Potential Explanation. Which should we choose? There is a natural temptation to plump for likeliness. After all, Inference to the Best Explanation is supposed to describe strong inductive arguments, and a strong inductive argument is one where the premises make the conclusion likely. But in fact this connection is too close and, as a consequence, choosing likeliness would push Inference to the Best Explanation towards triviality. We want a model of inductive inference to describe what principles we use to judge one inference more likely than another, so to say that we infer the likeliest explanation is not helpful. To put the point another way, we want our account of inference to give the *symptoms* of likeliness, the features an argument has that lead us to say that the premises make the conclusion likely. A model of Inference to the

Likeliest Explanation begs these questions. It would still have some content, since it suggests that inference is a matter of selection from among competitors and that inference is often inference to a cause. But for Inference to the Best Explanation to provide an illuminating account, it must say more than that we infer the likeliest cause (cf. Cartwright, 1983, p. 6). This gives us a second useful acid test for Inference to the Best Explanation. (The first one was that it must do better than the hypothetico-deductive model.) Inference to the Best Explanation is an advance only if it reveals more about inference than that it is often inference to the likeliest cause. It should show how likeliness is determined (at least in part) by explanatory considerations.

So the version of Inference to the Best Explanation we should consider is Inference to the Loveliest Potential Explanation. Here at least we have an attempt to account for epistemic value in terms of explanatory virtue. This version claims that the explanation that would, if true, provide the deepest understanding is the explanation that is likeliest to be true. Such an account suggests a really lovely explanation of our inferential practice itself, one that links the search for truth and the search for understanding in a fundamental way. Similar remarks apply to the notion of potential explanation, if we opt for the narrower live option characterization I favor. We want to give an account of the plausibility filter that determines the pool of potential explanations, and a deep version of Inference to the Best Explanation will give this characterization in explanatory terms: it will show how explanatory considerations determine plausibility.

The distinction between likeliness and loveliness is, I hope, reasonably clear. Nevertheless, it is easy to see why some philosophers may have conflated them. After all, if Inference to the Loveliest Explanation is a reasonable account, loveliness and likeliness will tend to go together, and indeed loveliness will be a guide to likeliness. Moreover, given the darkness of our 'inference box', we may be aware only of inferring what seems likeliest even if the mechanism actually works by assessing loveliness. Our awareness of what we are doing may not suggest the correct description. In any event, if there is a tendency to conflate the distinction, this helps to explain why Inference to the Best Explanation enjoys more popularity among philosophers than is justified by the arguments given to date in its favor. By implicitly construing the slogan simply as inference to the likeliest

explanation, it is rightly felt to apply to a wide range of inferences; by failing to notice the difference between this and the deep account, the triviality is suppressed. At the same time, the distinction between likeliness and loveliness, or one like it, is one that most people who were seriously tempted to develop the account would make, and this may help to explain why the temptation has been so widely resisted. Once one realizes that an interesting version requires an account of explanatory loveliness that is conceptually independent of likeliness, the weakness of our grasp on what makes one explanation lovelier than another is discouraging.

In practice, these two versions of Inference to the Best Explanation are probably ideal cases: a defensible version may well need to combine elements of each, accounting for likeliness only partially in explanatory terms. For example, one might construct a version where a non-explanatory notion of likeliness plays a role in restricting membership in the initial set of potential explanations, but where considerations of loveliness govern the choice from among the members of that set. Again, we may have to say that considerations of likeliness having nothing to do with explanation will, under various conditions, defeat a preference for loveliness. This may be the only way to account for the full impact of disconfirming evidence. So the distinction between likeliness and loveliness leaves us with considerable flexibility. But I think we may take it as a rule of thumb that the more we must appeal to likeliness analyzed in non-explanatory terms to produce a defensible version of Inference to the Best Explanation, the less interesting that model is. Conversely, the more use we can make of the explanatory virtues, the closer we will come to fulfilling the exciting promise of Inference to the Best Explanation, of showing how explanatory considerations are our guide to the truth.

We have now gone some way towards spelling out the slogan, by making the distinctions between potential and actual explanation and between the likeliest and the loveliest explanation. By seeing how easy it is to slide from loveliness to likeliness, we have also sensitized ourselves to the risk of trivializing the model by making it so flexible that it can be used to describe almost any form of inference. But there are also various respects in which the scope of Inference to the Best Explanation is greater than may initially appear. Two apparent and damaging consequences of

Inference to the Best Explanation are that only one explanation can be inferred from any set of data and that the only data that are relevant to a hypothesis are data the hypothesis explains. Both of these are, however, merely apparent consequences, on a reasonable version of Inference to the Best Explanation. The first is easily disposed of. Inference to the Best Explanation does not require that we infer only one explanation of the data, but that we infer only one of *competing* explanations. The data from a flight recorder recovered from the wreckage of an airplane crash may at once warrant explanatory inferences about the motion of the plane, atmospheric conditions at the time of the accident, malfunctions of equipment in the airplane, and the performance of the pilot, and not simply because different bits of information from the recorder will warrant different inferences, but because the same bits may be explained in many different but compatible ways. When I notice that my front door has been forced open, I may infer both that I have been robbed and that my deadbolt is not as force-resistant as the locksmith claimed.

Inference to the Best Explanation can also account for some of the ways evidence may be relevant to a hypothesis that does not explain it. The most obvious mechanism for this depends on a deductive consequence condition on inference. If I am entitled to infer a theory, I am also entitled to infer whatever follows deductively from that theory, or from that theory along with other things I reasonably believe (cf. Hempel, 1965, pp. 31–2). This is at least highly plausible: it would be a poor joke to say one is entitled to believe a theory but not its consequences. Suppose now that I use Inference to the Best Explanation to infer from the data to a high-level theory, and then use the consequence condition to deduce a lower-level hypothesis from it. There is now no reason to suppose that the lower-level theory will explain all of the original data that indirectly support it. Newton was entitled to infer his theory in part because it explained the result of various terrestrial experiments. This theory in turn entails laws of planetary orbit. Inference to the Best Explanation with its consequence condition counts those laws as supported by the terrestrial evidence, even though the laws do not explain that evidence. It is enough that the higher-level theory does so. The clearest cases of the consequence condition, however, are deduced predictions. What will happen does not explain what happened in the past, but a theory that entails the prediction may.

Since Inference to the Best Explanation will sometimes underwrite inferences to high-level theories, rich in deductive consequences, the consequence condition substantially increases the scope of the model. Even so, we may wish to broaden the scope of the condition to include 'explanatory consequences' as well as strictly deductive ones. Seeing the distinctive flash of light, I infer that I will hear thunder. The thunder does not explain the flash, but the electrical discharge does and would also explain why I will hear thunder. But the electrical discharge does not itself entail that I will hear thunder. It is not merely that there is more to the story, but that there is always the possibility of interference. I might go temporarily deaf, the lightning may be too far away, I might sneeze just at the wrong moment, and there are always other possibilities that I do not know about. Someone who favors deductive models will try to handle these possibilities by including extra premises, but this will always require an unspecifiable *ceteris paribus* clause. So in many cases, it may be more natural to allow 'Inference from the Best Explanation' (Harman, 1986, pp. 68–70). Noticing that it is extraordinarily cold this morning, I infer that my car will not start. The failure of my car to start would not explain the weather, but my inference is naturally described by saying that I infer that it will not start because the weather would provide a good explanation of this, even though it does not entail it. (The risk of interference also helps to explain why we are often more confident of inferences to an explanation than inferences from an explanation. When we start from the effect, we know that there was no effective interference.)

## ATTRACTIONS AND REPULSIONS

We have said enough to give some content to the idea of Inference to the Best Explanation. What the account now needs is some specific argument on its behalf, which I will begin to provide in the next chapter. First, however, it will be useful to compile a brief list of its general and *prima facie* advantages and disadvantages, some of which have already been mentioned, to prepare the ground for a more detailed assessment. Inference to the Best Explanation seems itself to be a relatively lovely explanation of our inductive practices. It gives a natural description of familiar aspects of our inferential procedures. The

simplest reason for this is that we are often aware that we are inferring an explanation of the evidence, but there is more to it than that. We are also often aware of making an inferential choice between competing explanations, and this typically works by means of the two-filter process my favored version of Inference to the Best Explanation describes. We begin by considering plausible candidate explanations and then try to find data that discriminate between them. The account reflects the fact that a hypothesis that is a reasonable inference in one competitive milieu may not be in another. An inference may be defeated when someone suggests a better alternative explanation, even though the evidence does not change. Inference to the Best Explanation also suggests that we assess candidate inferences by asking a subjunctive question: we ask how good the explanation *would* be if it were true. There seems to be no reason why an inferential engine has to work in this way. If induction really did work by simple extrapolation, it would not involve subjunctive assessment. We can also imagine an inductive technique that included selection from among competitors but did not involve the subjunctive process. We might simply select on the basis of some feature of the hypotheses that we directly assess. In fact, however, we do often make the inductive decision whether something is true by asking what would be the case if it were, rather than simply deciding which is the likeliest possibility. We construct various causal scenarios and consider what they would explain and how well. Why is my refrigerator not running? Perhaps the fuse has blown. Suppose it has; but then the kitchen clock should not run either, since the clock and refrigerator are on the same circuit. Is the clock running? By supposing for the moment that a candidate explanation is correct, we can work out what further evidence is relevant to our inference. The role of subjunctive reasoning is partially captured by the familiar observation about the 'priority of theory over data'. Induction does not, in general, work by first gathering all the relevant data and only then considering the hypotheses to which they apply, since we often need to entertain a hypothesis first in order to determine what evidence is relevant to it (Hempel, 1966, pp. 12–13). But the point about subjunctive evaluation is not only that explanatory hypotheses are needed to determine evidential relevance, but also a partial description of how that determination is made. (One of the attractions of the hypothetico-deductive

model is that it also captures this subjunctive aspect of our methods for assessing relevant evidence, since we determine what a hypothesis entails by asking what would have to be the case if the hypothesis were true.)

Although we often infer an explanation just because that is where our interests lie, Inference to the Best Explanation correctly suggests that explanatory inferences should be common even in cases where explaining is not our primary purpose. Even when our main interest is in accurate prediction or effective control, it is a striking feature of our inferential practice that we often make an 'explanatory detour'. If I want to know whether my car will start tomorrow, my best bet is to try to figure out why it sometimes hasn't started in the past. When Semmelweis wanted to control the outbreak of childbed fever in one of the maternity wards in the Vienna hospital where he worked, he proceeded by trying to explain why the women in that ward were contracting the disease, and especially the contrast between the women in that ward and the women in another ward in the same hospital, who rarely contracted it. The method of explanatory detour seems to be one of the sources of the great predictive and manipulative successes of many areas of science. In science, the detour often requires 'vertical' inference to explanations in terms of unobserved and often unobservable entities and processes, and Inference to the Best Explanation seems particularly well equipped to account for this process.

In addition to giving a natural description of these various features of our inferential practice, Inference to the Best Explanation has a number of more abstract attractions. The notion of explanatory loveliness, upon which an interesting version of Inference to the Best Explanation relies, should help to make sense of the common observation of scientists that broadly aesthetic considerations of theoretical elegance, simplicity, and unification are a guide to inference. More generally, as I have already mentioned, the account describes a deep connection between our inferential and explanatory behavior, one that accounts for the prevalence of explanatory inferences even in cases where our main interests lie elsewhere. As such, it also helps with one of the problems of justifying our explanatory practices, since it suggests that one of the reasons for our obsessive search for explanations is that this is a peculiarly effective way of discovering the structure of the world. The explicit point

of explaining is to understand *why* something is the case but, if Inference to the Best Explanation is correct, it is also our primary tool for discovering *what* is the case.

Another sort of advantage to the view that induction is Inference to the Best Explanation is that it avoids some of the objections to competing models of inductive inference or confirmation that we discussed in chapter one. One of the weaknesses of the simple Humean extrapolation model ('More of the Same') is that we are not always willing to extrapolate and, when we are, the account does not explain which of many possible extrapolations we actually choose. Inference to the Best Explanation does not always sanction extrapolation, since the best explanation for an observed pattern is not always that it is representative (Harman, 1965, pp. 90–1). Given my background knowledge, the hypothesis that I always win is not a good explanation of my early successes at the roulette wheel. Similarly, given a finite number of points on a graph marking the observed relations between two quantities, not every curve through those points is an equally good explanation of the data. One of the severe limitations of both the extrapolation view and the instantial model of confirmation is that they do not cover vertical inferences, where we infer from what we observe to something at a different level that is often unobservable. As we have seen, Inference to the Best Explanation does not have this limitation: it appears to give a univocal account of horizontal and vertical inferences, of inferences to what is observable and to the unobservable. The instantial model is also too permissive, since it generates the raven paradox. In chapter six, I will attempt to show how Inference to the Best Explanation helps to solve it.

Inference to the Best Explanation also seems a significant advance over the hypothetico-deductive model. First, while that model has very little to say about the 'context of discovery', the mechanisms by which we generate candidate hypotheses, the two-filter version of Inference to the Best Explanation suggests that explanatory considerations may apply to both the generation of candidates and the selection from among them. Second, since the deductive model is an account of confirmation rather than inference, it does not say when a hypothesis may actually be inferred. Inference to the Best Explanation does better here, since it brings in competition selection. Third, while the hypothetico-deductive model allows for vertical inference, it does not say

much about how 'high' the inference may legitimately go. It allows the evidence to confirm a hypothesis however distant from it, so long as auxiliary premises can be found linking the two. In the next chapter, I will argue that Inference to the Best Explanation rightly focuses the impact of evidence more selectively, so that only some hypotheses that can be made to entail the evidence are supported by it. Fourth, if the model limits auxiliary hypotheses to independently known truths, it is too restrictive, since evidence may support a hypothesis even though the evidence is not entailed by the hypothesis and those auxiliaries, and it may disconfirm a hypothesis without contradicting it. Inference to the Best Explanation allows for this sort of evidence, since explanation does not require deduction. Finally, Inference to the Best Explanation avoids several of the sources of over-permissiveness that are endemic to the deductive model. In addition to avoiding the raven paradox, Inference to the Best Explanation blocks confirmation by various non-explanatory deductions. One example is the confirmation of an arbitrary conjunction by a conjunct, since a conjunction does not explain its conjuncts. For these as well as for some of the other liabilities of the hypothetico-deductive model, a symptom of the relative advantages of Inference to the Best Explanation is that many of the problems of the hypothetico-deductive model of confirmation and of the deductive-nomological model of explanation 'cancel out': many of the counterexamples of the one are also counterexamples of the other. This suggests that the actual explanatory relation offers an improved guide to inference.

We have so far canvassed two sorts of advantages of Inference to the Best Explanation. The first is that it is itself a lovely explanation of various aspects of inference; the second is that it is better than the competition. The third and final sort of advantage I will mention is that, in addition to accounting for scientific and everyday inference, Inference to the Best Explanation has a number of distinctively philosophical applications. The first is that it accounts for its own discovery. In chapter one, I suggested that the task of describing our inductive behavior is itself a broadly inductive project, one of going from what we observe about our inferential practice to the mechanism that governs it. If this is right, a model of induction ought to apply to itself. Clearly the extrapolation and the instantial models do not do well on this criterion, since the inference is to the contents of a black box, the

sort of vertical inference those models do not sanction. Nor does the hypothetico-deductive model do much better, since it does not entail many observed features of our practice. It does not, for example, entail any of the inferences we actually make. Inference to the Best Explanation does much better on this score. The inference to an account of induction is an explanatory inference: we want to explain why we make the inferences we do. Our procedure has been to begin with a pool of plausible candidate explanations (the various models of induction we have canvassed) and then select the best. This is a process of competition selection which works in part by asking the subjunctive question of what sort of inferences we would make, if we used the various models. Moreover, if we do end up selecting Inference to the Best Explanation, it will not simply be because it seems the likeliest explanation, but because it has the features of unification, elegance, and simplicity that make it the loveliest explanation of our inductive behavior.

Another philosophical application of Inference to the Best Explanation is to the local justification of some of our inferential practices. For example, it is widely supposed that a theory is more strongly supported by successful predictions than by data that were known before the theory was constructed and which the theory was designed to accommodate. At the same time, the putative advantage of prediction over accommodation is controversial and puzzling, because the logical relationships between theory and data upon which inductive support is supposed to depend seem unaffected by the merely historical fact of when the data were observed. But there is a natural philosophical inference to the best explanation that seems to defend the epistemic distinction. When data are predicted, the best explanation for the fit between theory and data, it is claimed, is that the theory is true. When the data are accommodated, however, there is an alternative explanation of the fit, namely that the theory was designed just for that purpose. This explanation, which only applies in the case of accommodation, is better than the truth explanation, and so Inference to the Best Explanation shows why prediction is better than accommodation (cf. Horwich, 1982, p. 111). We will assess this argument in chapter eight.

Another example of the application of Inference to the Best Explanation to local philosophical justification is in connection with Thomas Kuhn's notorious discussion of 'incommensurability'

(Kuhn, 1970, esp. chs 9–10). According to him, there is no straightforward way of resolving scientific debates during times of 'scientific revolutions', because the disputants disagree about almost everything, including the evidence. This seems to block resolution by any appeal to a crucial experiment. On the traditional view of such experiments, they resolve theoretical disputes by providing evidence that simultaneously refutes one theory while supporting the other. Competing theories are found to make conflicting predictions about the outcome of some experiment; the experiment is performed and the winner is determined. But this account seems to depend on agreement about the outcomes of experiment, which Kuhn denies. These experiments can, however, be redescribed in terms of Inference to the Best Explanation in a way that does not assume shared observations. A crucial experiment now becomes two experiments, one for each theory. The outcome of the first experiment is explained by its theory, whereas the outcome of the second is not explained by the other theory, so we have some basis for a preference. Shared standards of explanation may thus compensate for observational disagreement: scientists should prefer the theory that best explains its proper data. There is, however, more to Kuhn's notion of incommensurability than disagreement over the data; in particular, there is also tacit disagreement over explanatory standards. But this may turn out to be another advantage of Inference to the Best Explanation. Insofar as Kuhn is right here, Inference to the Best Explanation will capture the resulting indeterminacy of scientific debate that is an actual feature of our inferential practices.

Another well-known philosophical application of Inference to the Best Explanation is to argue for various forms of realism. For example, as part of an answer to the Cartesian skeptic who asks how we can know that the world is not just a dream or that we are not just brains in vats, the realist may argue that we are entitled to believe in the external world since hypotheses that presuppose it provide the best explanation of our experiences. It is possible that it is all a dream, or that we are really brains in vats, but these are less good explanations of the course of our experiences than the ones we all believe, so we are rationally entitled to our belief in the external world. There is also a popular application of Inference to the Best Explanation to realism in the philosophy of science, which we have already briefly mentioned. The issue here

is whether scientific theories, particularly those that appeal to unobservables, are getting at the truth, whether they are providing an increasingly accurate representation of the world and its contents. There is an inference to the best explanation for this conclusion. In brief, later theories tend to have greater predictive success than those they replace, and the best explanation for this is that later theories are better descriptions of the world than earlier ones. We ought to infer scientific realism, because it is the best explanation of predictive progress. We will assess this argument in chapter nine.

Let us conclude this chapter with some of the bad news. First, several of the philosophical applications of Inference to the Best Explanation can be questioned. In the case of the argument for the advantages of prediction over accommodation, one may ask whether the 'accommodation explanation' really competes with the truth explanation (Horwich, 1982, pp. 112–16). If not, then as we saw in the last section, Inference to the Best Explanation does not require that we choose between them. Moreover, the assumption that they do compete, that explaining the fit between theory and accommodated data by appeal to the act of accommodation pre-empts explaining the fit by appeal to the truth of the theory, seems just to assume that accommodation does not provide support, and so to beg the question. As for the argument for realism about the external world, do our beliefs about the world really provide a better explanation than the dream hypothesis, or is it simply that this is the explanation we happen to prefer? Again, doesn't the inference to scientific realism as the best explanation for predictive success simply assume that inferences to the best explanation are guides to the truth about unobservables, which is just what an opponent of scientific realism would deny? A second sort of liability is the suspicion that Inference to the Best Explanation is still nothing more than Inference to the Likeliest Cause in fancy dress, and so fails to account for the symptoms of likeliness. Third, insofar as there is a concept of explanatory loveliness that is conceptually distinct from likeliness, one may question whether this is a suitable criterion of inference. On the one hand, there is what we may call 'Hungerford's objection', in honor of the author of the line, 'Beauty is in the eye of the beholder'. Perhaps explanatory loveliness is too subjective and interest relative to give an account of inference that reflects the objective features of inductive warrant.

On the other hand, supposing that loveliness is as objective as inference, we have 'Voltaire's objection'. What reason is there to believe that the explanation that would be loveliest, if it were true, is also the explanation that is most likely to be true? Why should we believe that we inhabit the loveliest of all possible worlds? As we saw in the last section, Inference to the Best Explanation requires that we work with a notion of potential explanation that does not carry a truth requirement. Once we have removed truth from explanation, however, it is not clear how we get it back again (cf. Cartwright, 1983, pp. 89–91). Lastly, perhaps most importantly, it will be claimed that Inference to the Best Explanation is only as good as our account of explanatory loveliness, and this account is nonexistent. In the next chapter, I begin to meet this objection.

# 5

# CONTRASTIVE INFERENCE

## A CASE STUDY

In this chapter and the next two, I will consider some of the prospects of Inference to the Best Explanation as a solution to the descriptive problem of inductive inference. We want to determine how illuminating that account is as a description, or a partial description, of the mechanism inside the cognitive black box that governs our inductive practices. To do this, we need to show how explanatory considerations are a guide to inference, how loveliness helps to determine likeliness. In particular, we want to see whether the model can meet the two central challenges from the last chapter, to show that inferences to the best explanation are more than inferences to the likeliest cause, and to show that Inference to the Best Explanation marks an advance over the simple hypothetico-deductive model.

As I have stressed, the main difficulty standing in the way of this project is our poor understanding of what makes one explanation lovelier than another. Little has been written on this subject, perhaps because it has proven so difficult even to say what makes something an explanation. How can we hope to determine what makes one explanation better than another, if we can't even agree about what distinguishes explanations of any quality from things that are not explanations at all? Moreover, most of what has been written about explanatory loveliness has focused on the interest relativity of explanation, which seems to bring out pragmatic and subjective factors that are too variable to provide a suitably objective measure of inductive warrant.

Yet the situation is not hopeless. My analysis of contrastive explanation in chapter three will help. There I argued that pheno-