Chapter 5

# Contrastive inference

## A case study

In this chapter and the next four, 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 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 or confirmation; 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, a major challenge facing 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 cannot even agree about what distinguishes explanations of any quality from something that is not an explanation 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 3 will help. There I argued that phenomena we explain often have a contrastive fact–foil structure, and that the foil helps to select the part of the causal history of the fact that provides a good explanation by means of a mechanism of causal triangulation. According to my Difference Condition, to explain why P rather than Q, we need a causal difference between P and not-Q, consisting of a cause of P and the absence of a corresponding event in the case of not-Q. Thus we can explain why Jones rather than Smith contracted paresis by pointing out that only Jones had syphilis, since this is to point out a causal difference between the two men, even though most people with syphilis do not get paresis. This account of contrastive explanation shows how what counts as a good explanation depends on interests, since interests determine the choice of foil, and a cause that marks a difference for one foil will not generally do so for another. Jones's syphilis would not explain why he rather than Doe (who also had syphilis) contracted paresis; here the explanation might be instead that only Jones left his syphilis untreated. In this respect, then, what counts as a lovely explanation of P depends on one's interests, but this cashes out into the question of whether the cited cause provides any explanation at all of the contrast that expresses a particular interest.

The sensitivity of explanation to choice of foils captures much of what has been said about interest relativity, and it also shows that these factors are not strongly subjective in a way that would make them irrelevant to inference. An account of the interest relativity of explanation would be strongly subjective if it showed that what counts as a good explanation depends on the tastes of the audience rather than the causal structure of the world. Examples that would threaten the idea of Inference to the Best Explanation would be cases where people favor incompatible explanations of the same phenomenon, even though their evidence and their inferential inclinations are the same. It is no threat to the objectivity of explanation that different people should be interested in explaining different phenomena, and it is obvious that a good explanation of one phenomenon is not usually a good explanation of another. A contrastive analysis of explanation supports only this innocuous form of relativity, if we construe the phenomena themselves as contrastive, so that a change in foil yields a different phenomenon. Moreover, my analysis of contrastive explanation shows that a change in foil helps to select a different part of the same causal history. Differences in interest require different but compatible explanations, which does not bring in strong subjectivity. And this much interest relativity is also something any reasonable account of inference must acknowledge: different people may all reasonably infer different things from shared evidence, depending on their inferential interests, when the inferences are compatible.

So my account of contrastive explanation helps to defuse the objection to Inference to the Loveliest Explanation that loveliness is hopelessly subjective. (We will return to this issue in later chapters.) It also provides the core of a positive account of one way that explanatory considerations can serve as a guide to inference. The reason for this is the structural similarity between the Difference Condition and Mill's Method of Difference. According to Mill, we find the cause of a fact in some prior difference between a case where the fact occurs and an otherwise similar case where it does not. Mill's central mechanism for inferring the likeliest cause is almost the same as the mechanism of causal triangulation that helps to determine the loveliest explanation. This near-isomorphism provides an important argument in favor of Inference to the Best Explanation, since it shows that a criterion we use to evaluate the quality of potential explanations is the same as one we use to infer causes. By inferring something that would provide a good explanation of the contrast if it were a cause, we are led to infer something that is likely to be a cause. Returning to poor Jones, we may find that his condition, taken alone, points to no particular explanation. But if we try instead to explain why Jones rather than Smith contracted paresis, we will be led, by means of the Difference Condition, to look for some possibly relevant difference in the medical histories of the two men. Thus we may infer that Jones's syphilis was a cause of his paresis, since this is an explanatory difference. And this is just where Mill's method would take us, if syphilis was the only possibly relevant difference. Moreover, our explanation and our inference will both change if we change the foil. If we ask why Jones rather than Doe contracted paresis, we will be led to explain this contrast by appeal to Doe's treatment. By varying the foil, we change the best explanation, and this leads us to different but compatible inductive inferences, taking us to different stages of Jones's medical history.

By considering inferences to contrastive explanations, we go some way towards meeting the challenge that Inference to the Best Explanation is nothing more than Inference to the Likeliest Cause, where likeliness is judged on some basis entirely independent of explanatory considerations. Since looking for residual differences in similar histories of fact and foil is a good way of determining a likely cause, as Mill taught us, and contrastive explanation depends on just such differences, looking for potential contrastive explanations can be a guide to causal inference. Given contrastive data, the search for explanation is an effective way of determining just what sort of causal hypotheses the evidence supports. This procedure focuses our inferences, by eliminating putative causes that are in the shared part of antecedents of fact and foil. These antecedents may well be causally relevant, but the fact that they would not explain the contrast shows that the contrast does not (at least by itself) provide evidence that they are causes. This version of Inference to the Best Explanation thus sheds some light on the context of discovery, since the requirement that a potential explanation cite a difference severely restricts the class of candidate hypotheses. It also brings out one role of background knowledge in inference in a natural way, since our judgment of which antecedents are shared, a judgment essential to the application of the method, will depend on such knowledge.

I also want to argue, in this chapter and the next, that Inference to the Best Contrastive Explanation helps to meet the second challenge, to show that the model is better than simple hypothetico-deductivism. It marks an improvement both where the deductive model is too strict, neglecting evidential relevance in cases where there is no appropriate deductive connection between hypothesis and data, and where it is too lenient, registering support where there is none to be had as, for example, revealed by the raven paradox. Inference to the Best Explanation does better in the first case because, as the analysis of contrastive explanation shows, explanatory causes need not be sufficient for their effects, so the fact that a hypothesis would explain a contrast may provide some reason to believe the hypothesis, even though the hypothesis does not entail the data. It does better in the second case because, while some contrapositive instances (e.g. non-black non-ravens) do support a hypothesis, not all do, and the requirement of shared antecedents helps to determine which do and which do not. The structural similarity between the Method of Difference and contrastive explanation that I will exploit in these chapters will also eventually raise the question of why Inference to the Best Explanation is an improvement on Mill's methods, a question I will address in chapter 8.

To develop these arguments and, more generally, to show just how inferences to contrastive explanations work, it is useful to consider a simple but actual scientific example in some detail. The example I have chosen is Ignaz Semmelweis's research from 1844–8 on childbed fever, inspired by Carl Hempel's well-known and characteristically clear discussion (1966: 3–8) and Semmelweis's own account of his work (1860). Semmelweis wanted to find the cause of this often fatal disease, which was contracted by many of the women who gave birth in the Viennese hospital in which he did his research. Semmelweis's central datum was that a much higher percentage of the women in the First Maternity Division of the hospital contracted the disease than in the adjacent Second Division, and Semmelweis sought to explain this difference. The hypotheses he considered fell into three types. In the first were hypotheses that did not mark differences between the divisions, and so were rejected. Thus, the theory of 'epidemic influences' descending over entire districts did not explain why more women should die in one division than another; nor did it explain why the mortality among Viennese women who gave birth at home or on the way to the hospital was lower than in the First Division. Similarly, the hypotheses that the fever is caused by overcrowding, by diet or by general care were rejected because these factors did not mark a difference between the divisions.

One striking difference between the two divisions was that medical students only used the First Division for their obstetrical training, while midwives received their training in the Second Division. This suggested the hypothesis that the high rate of fever in the First Division was caused by injuries due to rough examination by the medical students. Semmelweis rejected the rough examination hypothesis on the grounds that midwives performed their examinations in more or less the same way, and that the injuries due to childbirth are in any case greater than those due to rough examination.

The second type of hypotheses were those that did mark a difference between the divisions, but where eliminating the difference in putative cause did not affect the difference in mortality. A priest delivering last sacrament to a dying woman had to pass through the First Division to get to the sickroom where dying women were kept, but not through the Second Division. This suggested that the psychological influence of seeing the priest might explain the difference, but Semmelweis ruled this out by arranging for the priest not to be seen by the women in the First Division either and finding that this did not affect the mortality rates. Again, women in the First Division delivered lying on their backs, while women in the Second delivered on their sides, but when Semmelweis arranged for all women to deliver on their sides, the mortality remained the same.

The last type of hypothesis that Semmelweis considered is one that marked a difference between the divisions, and where eliminating this difference also eliminated the difference in mortality. Kolletschka, one of Semmelweis's colleagues, received a puncture wound in his finger during an autopsy, and died from an illness with symptoms like those of childbed fever. This led Semmelweis to infer that Kolletschka's death was due to the 'cadaveric matter' that the wound introduced into his blood stream, and Semmelweis then hypothesized that the same explanation might account for the deaths in the First Division, since medical students performed their examinations directly after performing autopsies, and midwives did not perform autopsies at all. Similarly, the cadaveric hypothesis would explain why women who delivered outside the hospital had a lower mortality from childbed fever, since they were not examined. Semmelweis had the medical students disinfect their hands before examination, and the mortality rate in the First Division went down to the same low level as that in the Second Division. Here at last was a difference that made a difference, and Semmelweis inferred the cadaveric hypothesis.

This case is a gold mine for inferences to the best contrastive explanation. Let us begin by considering Semmelweis's strategy for each of the three groups of hypotheses: those of no difference, of irrelevant differences and of relevant differences. Semmelweis's rejection of the hypotheses in the first group – epidemic influences, overcrowding, general care, diet and rough examination – show how Inference to the Best Explanation can account for negative evidence. These hypotheses are rejected on the grounds that, though they are compatible with the evidence, they would not explain the contrast between the divisions. Epidemic influences, for example, still might possibly be part of the causal history of the deaths in the First Division, say because the presence of these influences is a necessary condition for any case of childbed fever. And nobody who endorsed the epidemic hypothesis would have claimed that the influences are sufficient for the fever, since it was common knowledge that not all mothers in the district contracted childbed fever. Still, Semmelweis took the fact that the hypotheses in the first group would not explain the contrast between the divisions or the contrast between the First Division and mothers who gave birth outside the hospital to be evidence against them.

Semmelweis also used a complementary technique for discrediting the explanations in the first group that is naturally described in terms of Inference to the Best Explanation, when he argued against the epidemic hypothesis on the grounds that the mortality rate for births outside the hospital was lower than in the First Division. What he has done is to change the foil, and point out that the hypothesis also fails to explain this new contrast. It explains neither why mothers get fever in the First Division rather than in the Second, nor why mothers get fever in the First Division rather than outside the hospital. Similarly, when Semmelweis argued against the rough examination hypothesis on the grounds that childbirth is rougher on the mother than any examination, he pointed out not only that it fails to explain why there is fever in the First Division rather than in the Second, but also why there is fever in the First Division rather than among other mothers generally. New foils provide new evidence, in these cases additional evidence against the putative explanations.

The mere fact that the hypotheses in the first group did not explain some evidence can not, however, account for Semmelweis's negative judgment. No hypothesis explains every observation, and most evidence that is not explained by a hypothesis is simply irrelevant to it. But Semmelweis's observation that the hypotheses do not explain the contrast in mortality between the divisions seems to count against those hypotheses in a way that, say, the observation that those hypotheses would not explain why the women in the First Division were wealthier than those in the Second Division (if they were) would not. Of course, since Semmelweis was interested in reducing the incidence of childbed fever, he was naturally more interested in an explanation of the contrast in mortality than in an explanation of the contrast in wealth, but this does not show why the failure of the hypotheses to explain the first contrast counts against them. This poses a general puzzle for Inference to the Best Explanation: how can that account distinguish negative evidence from irrelevant evidence, when the evidence is logically consistent with the hypothesis?

One straightforward mechanism is rival support. In some cases, evidence counts against one hypothesis by improving the explanatory power of a competitor. The fact that the mortality in the First Division went down when the medical students disinfected their hands before examination supports the cadaveric matter hypothesis, and so indirectly counts against all the hypotheses inconsistent with it that cannot explain this contrast. But this mechanism of disconfirming an explanation by supporting a rival does not seem to account for Semmelweis's rejection of the hypotheses in the first group, since at that stage of his inquiry he had not yet produced an alternative account.

Part of the answer to this puzzle about the difference in the epistemic relevance of a contrast in mortality and a contrast in wealth is that the rejected hypotheses would have enjoyed some support from the fact of mortality but not from the fact of wealth. The epidemic hypothesis, for example, was not Semmelweis's invention, but a popular explanation at the time of his research. Its acceptance presumably depended on the fact that it seemed to provide an explanation, if a weak one, for the non-contrastive observations of the occurrence of childbed fever. In the absence of a stronger and competing explanation, this support might have seemed good enough to justify the inference. But by pointing out that the hypothesis does not explain the contrast between the divisions, Semmelweis undermines this support. On the other hand, the epidemic hypothesis never explained and so was never supported by observations about the wealth of the victims of childbed fever, so its failure to explain why the women in the First Division were wealthier than those in the Second Division would not take away any support it had hitherto enjoyed.

On this view, the observation that the hypotheses in the first group do not explain the contrast in mortality and the observation that they do not explain the contrast in wealth are alike in that they both show that these data do not support the hypothesis. The difference in impact only appears when we take into account that only evidence about mortality had been supposed to support the hypothesis, so only in this case is there a net loss of support. This view seems to me to be correct as far as it goes, but it leaves a difficult question. Why, exactly, does the failure to explain the contrast in mortality undermine prior support for hypotheses in the first group? Those hypotheses would still give some sort of explanation for the cases of the fever in the hospital, even if they would not explain the contrast between the divisions. Consider a different example. Suppose that we had two wards of patients who suffer from syphilis and discovered that many more of them in one ward contracted paresis than in the other. The hypothesis that syphilis is a necessary cause of paresis would not explain this contrast, but this would not, I think, lead us to abandon the hypothesis on the grounds that its support had been undermined. Instead, we would continue to accept it and look for some further and complementary explanation for the difference between the wards, say in terms of a difference in the treatments provided. Why, then, is Semmelweis's case any different?

The difference must lie in the relative weakness of the initial evidence in support of the hypotheses in the first group. If the only evidence in favor of the epidemic hypothesis is the presence of childbed fever, the contrast in mortality does undermine the hypothesis, because it suggests that the correct explanation of the contrast will show that epidemic influences have nothing to do with fever. If, on the other hand, the epidemic hypothesis would also explain why there were outbreaks of fever at some times rather than others, or in some hospitals rather than others, even though these cases seemed similar in all other plausibly relevant respects, then we would be inclined to hold on to that hypothesis and look for a complementary explanation of the contrast between the divisions. In the case of syphilis and paresis, we presumably have extensive evidence that there are no known cases of paresis not preceded by syphilis. The syphilis hypothesis not only would explain why those with paresis have it, but also the many contrasts between people with paresis and those without it. This leads us to say that the correct explanation of the contrast between the wards is more likely to complement the syphilis hypothesis than to replace it.

If this account is along the right lines, then the strength of the disconfirmation provided by the failure to explain a contrast depends on how likely it seems that the correct explanation of the contrast will pre-empt the original hypothesis. This explains our different reaction to the wealth case. We may have no idea why the women in the First Division are wealthier than those in the second, but it seems most unlikely that the reason for this will pre-empt the hypotheses of the first group. When we judge that pre-emption is likely, we are in effect betting that the best explanation of the contrast will either contradict the original hypothesis or show it to be unnecessary, and so that the evidence that originally supported it will instead support a competitor. So the mechanism here turns out to be an attenuated version of disconfirmation by rival support after all. The inability of the hypotheses in the first group to explain the contrast between the divisions and the contrast between the First Division and births outside the hospital disconfirms those hypotheses because, although the contrastive data do not yet support a competing explanation, since none has yet been formulated, Semmelweis judged that the best explanation of those contrasts would turn out to be a competing rather than a complementary account. This judgment can itself be construed as an overarching inference to the best explanation. If we reject the hypotheses in the first group because they fail to explain the contrasts, this is because we regard the conjecture that the hypotheses are wrong to be a better explanation of the failures than that they are merely incomplete. Judgments of this sort are speculative, and we may in the end find ourselves inferring an explanation of the contrasts that is compatible with the hypotheses in the first group, but insofar as we do take their explanatory failures to count against them, I think it must be because we do make these judgments.

On this view, given a hypothesis about the etiology of a fact, and faced with the failure of that hypothesis to explain a contrast between that fact and a similar foil, the scientist must choose between the overarching explanations that the failure is due to incompleteness or that it is due to incorrectness. Semmelweis's rejections of the hypotheses in the first group are examples of choosing the incorrectness explanation. It is further corroboration of the claim that these choices must be made that we cannot make sense of Semmelweis's research without supposing that he also sometimes inferred incompleteness. For while the cadaveric hypothesis had conspicuous success in explaining the contrast between the divisions, it fails to explain other contrasts that formed part of Semmelweis's evidence. For example, it does not explain why some women in the Second Division contracted childbed fever while others in that division did not, since none of the midwives who performed the deliveries in that division performed autopsies. Similarly, the cadaveric hypothesis does not explain why some women who had street births on their way to the hospital contracted the fever, since those women were rarely examined by either medics or midwives after they arrived. Consequently, if we take it that Semmelweis nevertheless had good reason to believe that infection by cadaveric matter was a cause of childbed fever, it can only be because he reasonably inferred that the best explanation of these explanatory failures was only that the cadaveric hypothesis is incomplete, not the only cause of the fever, rather than that it is incorrect. These cases also show that we cannot in general avoid the speculative judgment by waiting until we actually produce an explanation for all the known relevant contrasts, since in many cases this would postpone inference indefinitely.

Let us turn now to the two hypotheses of the second group, concerning the priest and delivery position. Unlike the hypotheses of the first group, these did mark differences between the divisions and so might explain the contrast in mortality. The priest bearing the last sacrament only passed through the First Division, and only in that division did mothers deliver on their backs. Since these factors were under Semmelweis's control, he tested these hypotheses in the obvious way, by seeing whether the contrast in mortality between the divisions remained when these differences were eliminated. Since that contrast remained, even when the priest was removed from the scene and when the mothers in both divisions delivered on their sides, these hypotheses could no longer be held to explain the original contrast.

This technique of testing a putative cause by seeing whether the effect remains when it is removed is widely employed. Semmelweis could have used it even without the contrast between the divisions, and it is worth seeing how a contrastive analysis could account for this. Suppose that all the mothers in the hospital delivered on their backs, and Semmelweis tested the hypothesis that this delivery position is a cause of childbed fever by switching positions. He might have only done this for some of the women, using the remainder as a control. In this case, the two groups would have provided a potential contrast. If a smaller percentage of the women who delivered on their sides contracted childbed fever, the delivery hypothesis would have explained and so been supported by this contrast. And even if Semmelweis had switched all the mothers, he would have had a potential diachronic contrast, by comparing the incidence of fever before and after the switch. In either case, a contrast would have supported the explanatory inference. In fact, however, these procedures would not have produced a contrast, since delivery position is irrelevant to childbed fever. This absence of contrast would not disprove the delivery hypothesis. Delivering on the back might still be a cause of fever, but there might be some obscure alternate cause that came into play when the delivery position was switched. But the absence of the contrast certainly would disconfirm the delivery hypothesis. The reason for this is the same as in the case of the epidemic hypothesis: the likeliness of a better, pre-emptive explanation. Even if Semmelweis did not have an alternative explanation for the cases of fever, there must be another explanation in the cases of side delivery, and it is likely that this explanation will show that back delivery is irrelevant even when it occurs. As in the case of the hypotheses in the first group, when we take an explanatory failure to count against a hypothesis, even when we do not have an alternative explanation, this is because we infer that the falsity of the hypothesis is a better explanation for its explanatory failure than its incompleteness.

This leaves us with Semmelweis's final hypothesis, that the difference in mortality is explained by the cadaveric matter that the medical students introduced into the First Division. Here too we have an overarching explanation in play. Semmelweis had already conjectured that the difference in mortality was somehow explained by the fact that mothers were attended by medical students in the First Division and by midwives in the Second Division. This had initially suggested the hypothesis that the rough examinations given by the medical students was the cause, but this neither explained the contrast between the divisions nor the contrast between the mothers in the First Division and mothers generally, who suffer more from labor and childbirth than from any examination. The cadaveric hypothesis is another attempt to explain the difference between the divisions under the overarching hypothesis that the contrast is due to the difference between medical students and midwives. In addition to explaining the difference between divisions, this hypothesis would explain Kolletschka's illness, as well as the difference between the First Division and births outside the hospital.

Finally, Semmelweis tested this explanation by eliminating the cadaveric matter with disinfectant and finding that this eliminated the difference in the mortality between the divisions. This too can be seen as the inference to a contrastive explanation for a new contrast, where now the difference that is explained is not the simple difference in mortality between the divisions, but the diachronic contrast between the initial presence of that difference and its subsequent absence. The best explanation of the fact that removing the cadaveric matter is followed by the elimination of the difference in mortality is that it was the cadaveric matter that was responsible for that difference. By construing Semmelweis's evidence as a diachronic contrast, we bring out the important point that the comparative data have a special probative force that we would miss if we simply treated them as two separate confirmations of Semmelweis's hypothesis.

Semmelweis's research into the causes of childbed fever brings out many of the virtues of Inference to the Best Explanation when that account is tied to a model of contrastive explanation. In particular, it shows how explanatory considerations focus and direct inquiry. Semmelweis's work shows how the strategy of considering potential contrastive explanations focuses inquiry, even when the ultimate goal is not simply an explanation. Semmelweis's primary interest was to eliminate or at least reduce the cases of childbed fever, but he nevertheless posed an explanatory question: Why do women contract childbed fever? His initial ignorance was such, however, that simply asking why those with the fever have it did not generate a useful set of hypotheses. So he focused his inquiry by asking contrastive why-questions. His choice of the Second Division as foil was natural because it provided a case where the effect is absent yet the causal histories are very similar. By asking why the contrast obtains, Semmelweis focused his search for explanatory hypotheses on the remaining differences. This strategy is widely applicable. If we want to find out why some phenomenon occurs, the class of possible causes is often too big for the process of Inference to the Best Explanation to get a handle on. If, however, we are lucky or clever enough to find or produce a contrast where fact and foil have similar histories, most potential explanations are immediately 'cancelled out' and we have a manageable and directed research program. The contrast will be particularly useful if, as in Semmelweis's case, in addition to meeting the requirement of shared history, it is also a contrast that various available hypotheses will not explain. Usually, this will still leave more than one hypothesis in the field, but then further observation and experiment may produce new contrasts that leave only one explanation. This shows how the interest relativity of explanation is at the service of inference. By tailoring his explanatory interests (and his observational and experimental procedures) to contrasts that would help to discriminate between competing hypotheses, Semmelweis was able to judge which hypothesis would provide the best overall explanation of the wide variety of contrasts (and absences of contrast) he observed, and so to judge which hypothesis he ought to infer. Semmelweis's inferential interests determined his explanatory interests, and the best explanation then determined his inference.

Before assessing the prospects for a hypothetico-deductive analysis of Semmelweis's research, it is worth mentioning that while I have followed Hempel's sensible selection from among the arguments, Semmelweis's own presentation contains many more. Indeed he provides a kind of orgy of arguments from explanatory power that can only warm the heart of defenders of Inference to the Best Explanation. As Semmelweis himself put it, 'As soon as one knows that childbed fever rises from decaying matter which is conveyed from external sources, explanations are easy' (1860: 156). In his introduction to the translation of Semmelweis's work, K. Codell Carter gives a good sense of this remarkable explanatory range:

> Semmelweis drew from his account explanations for dozens of facts that had been recorded but never explained. To choose only a few examples, Semmelweis explained why infants never died from [childbed] fever while their mothers remained healthy, why the mortality rates of infants changed in certain ways, why women who delivered on the way to hospital or who delivered prematurely had a lower mortality rate, why the disease often appeared in particular patterns among patients, why the mortality rate was different in the two clinics and why it had changed in certain ways through history, why infections were rare during pregnancy or after delivery, why the disease appeared to be contagious, why it exhibited seasonal patterns, why the disease was concentrated in teaching hospitals, why some non-teaching hospitals had much lower mortality rate that others, and why the disease appeared with different frequencies in different countries and in different historical periods. (Semmelweis 1860: 39–40)

Semmelweis's hypothesis would explain all these things and his account of his work is a sustained argument that we should accept that hypothesis precisely because of this explanatory power and because of the failure of competing hypotheses to provide equally good explanations.

## Explanation and deduction

Semmelweis's research is a striking illustration of inferences to the best explanation in action, and of the way they often exploit contrastive data. It is also Hempel's paradigm of the hypothetico-deductive method. So this case is particularly well suited for a comparison of the virtues of Inference to the Best Explanation and the deductive model. It shows, I will suggest, that Inference to the Best Explanation is better than hypothetico-deductivism.

Consider first the context of discovery. Semmelweis's use of contrasts and prior differences to help generate a list of candidate hypotheses illustrates one of the ways Inference to the Best Explanation elucidates the context of discovery, a central feature of our inductive practice neglected by the hypothetico-deductive model. The main reason for this neglect is easy to see. Hypothetico-deductivists emphasize the hopelessness of narrow inductivism, the view that scientists ought to proceed by first gathering all the relevant data without theoretical preconception and then using some inductive algorithm to infer from those data to the hypothesis they best support. Scientists never have all the relevant data, they often cannot tell whether or not a datum is relevant without theoretical guidance, and there is no general algorithm that could take them from data to a hypothesis that refers to entities and processes not mentioned in the data (Hempel 1966: 10–18). The hypothetico-deductive alternative is that, while scientists never have all the data, they can at least determine relevance if the hypothesis comes first.

Given a conjectural hypothesis, they know to look for data that either can be deduced from it or would contradict it. The cost of this account is that we are left in the dark about the source of the hypotheses themselves. According to Hempel, scientists need to be familiar with the current state of research, and the hypotheses they generate should be consistent with the available evidence but, in the end, generating good hypotheses is a matter of 'happy guesses' (1966: 15).

The hypothetico-deductivist must be right in claiming that there are no universally shared mechanical rules that generate a unique hypothesis from any given pool of data since, among other things, different scientists generate different hypotheses, even when they are working with the same data. Nevertheless, this 'narrow hypothetico-deductivist' conception of inquiry badly distorts the process of scientific invention. Most hypotheses consistent with the data are non-starters, and the use of contrastive evidence and explanatory inference is one way the field is narrowed. In advance of an explanation for some effect, we know to look for a foil with a similar history. If we find one, this sharply constrains the class of hypotheses that are worth testing. A reasonable conjecture must provide a potential explanation of the contrast, and most hypotheses that are consistent with the data will not provide this. (For hypotheses that traffic in unobservables, the restriction to potential contrastive explanations still leaves a lot of play: we will consider further ways the class of candidate hypotheses is restricted in chapters 8 and 9.)

The slogan 'Inference to the Best Explanation' may itself bring to mind an excessively passive picture of scientific inquiry, suggesting perhaps that we simply infer whatever seems the best explanation of the data we happen to have. But the Semmelweis example shows that the account, properly construed, allows for the feedback between the processes of hypothesis formation and data acquisition that characterizes actual inquiry. Contrastive data suggest explanatory hypotheses, and these hypotheses in turn suggest manipulations and controlled experiments that may reveal new contrasts that help to determine which of the candidates is the best explanation. This is one of the reasons the subjunctive element in Inference to the Best Explanation is important. By considering what sort of explanation the hypothesis would provide, if it were true, we assess not only how good an explanation it would be, but also what as yet unobserved contrasts it would explain, and this directs future observation and experiment. Semmelweis's research also shows that Inference to the Best Explanation is well suited to describe the role of overarching hypotheses in directing inquiry. Semmelweis's path to his cadaveric hypothesis is guided by his prior conjecture that the contrast in mortalities between the divisions is somehow due to the fact that deliveries are performed by medical students in the First Division, but by midwives in the Second Division. He then searches for ways of fleshing out this explanation and for the data that would test various proposals. Again, I have suggested that we can understand Semmelweis's rejection of the priest and the birth position hypotheses in terms of an inference to a negative explanation. The best explanation for the observed fact that eliminating these differences between the divisions did not affect mortality is that the mortality had a different cause. In both cases, the intermediate explanations focus research, either by marking the causal region within which the final explanation is likely to be found, or by showing that a certain region is unlikely to include the cause Semmelweis is trying to find.

The hypothetico-deductive model emphasizes the priority of theory as a guide to observation and experiment, at the cost of neglecting the sources of theory. I want now to argue that the model also fails to give a good account of the way scientists decide which observations and experiments are worth making. According to the deductive model, scientists should check the observable consequences of their theoretical conjectures, or of their theoretical systems, consisting of the conjunction of theories and suitable auxiliary statements. As we will see below, this account is too restrictive, since there are relevant data not entailed by the theoretical system. As we have already seen, it is also too permissive, since most consequences are not worth checking. Any hypothesis entails the disjunction of itself and any observational claim whatever, but establishing the truth of such a disjunction by checking the observable disjunct rarely has any bearing on the truth of the hypothesis. The contrastive account of Inference to the Best Explanation is more informative, since it suggests Semmelweis's strategy of looking for observable contrasts that distinguish one causal hypothesis from competing explanations.

Even if we take both Semmelweis's hypotheses and his data as given, the hypothetico-deductive model gives a relatively poor account of their relevance to each other. This is particularly clear in the case of negative evidence. According to the deductive model, evidence disconfirms a hypothesis just in case the evidence either contradicts the hypothesis or contradicts the conjunction of the hypothesis and suitable auxiliary statements. None of the hypotheses Semmelweis rejects contradicts his data outright. For example, the epidemic hypothesis does not contradict the observed contrast in mortality between the divisions. Proponents of the epidemic hypothesis would have acknowledged that, like any other epidemic, not everyone who is exposed to the influence succumbs to the fever. They realized that not all mothers contract childbed fever, but rightly held that this did not refute their hypothesis, which was that the epidemic influence was a cause of the fever in those mothers that did contract it. So the hypothesis does not entail that the mortality in the two divisions is the same. Similarly, the delivery position hypothesis does not entail that the mortality in the two divisions is different when the birth positions are different; nor does it entail that the mortality will be the same when the positions are the same. Even if back delivery is a cause of childbed fever, the mortality in the

Second Division could have been as high as in the First, because the fever might have had other causes there. Similarly, the possibility of additional causes shows that back delivery could be a cause of fever even though the mortality in the First Division is lower than in the Second Division where all the mothers deliver on their sides. The situation is the same for all the other hypotheses Semmelweis rejects.

What does Hempel say about this? He finds a logical conflict, but in the wrong place. According to him, the hypotheses that appealed to overcrowding, diet or general care were rejected because the claims that the difference in mortality between the divisions was due to such differences 'conflict with readily observable facts' (1966: 6). The claim that, for example, the difference in mortality is due to a difference in diet is incompatible with the observation that there is no difference in diet. These are clearly cases of logical incompatibility, but they are not the ones Hempel needs: the claims that are incompatible with observation are not the general hypotheses Semmelweis rejects. Like the cadaveric hypothesis he eventually accepts, the hypotheses of overcrowding, diet and care are surely general conjectures about causes of childbed fever, not specific claims about the differences between the divisions. But the hypotheses that overcrowding, diet or general care is a cause of childbed fever is logically compatible with everything Semmelweis observes.

The hypothetico-deductivist must claim that hypotheses are rejected because, although they are compatible with the data, each of them, when conjoined with suitable auxiliary statements, is not. But what could such statements be? Each hypothesis must have a set of auxiliaries that allows the deduction that the mortality in the divisions is the same, which contradicts the data. The auxiliaries need not be known to be true, but they need to be specified. This, however, cannot be done. The proponent of the epidemic hypothesis, for example, does not know what additional factors determine just who succumbs to the influence, so he cannot say how the divisions must be similar in order for it to follow that the mortality should be the same. Similarly, Semmelweis knew from the beginning that back delivery cannot be necessary for childbed fever, since there are cases of fever in the Second Division where all the women delivered on their sides, but he cannot specify what all the other relevant factors ought to be. The best the hypothetico-deductivist can do, then, is to rely on *ceteris paribus* auxiliaries. If fever is caused by epidemic influence, or by back delivery, and everything else 'is equal', the mortality in the divisions ought to be the same. This, however, does not provide a useful analysis of the situation. Any proponent of the rejected hypotheses will reasonably claim that precisely what the contrast between the divisions shows is that not everything is equal. This shows that there is more to be said about the etiology of childbed fever, but it does not show why we should reject any of the hypotheses that Semmelweis does reject. Semmelweis's observations show that none of these hypotheses would explain the contrasts, but they do not show that the hypotheses are false, on hypothetico-deductive grounds.

There is another objection to the *ceteris paribus* approach, and indeed to any other scheme that would generate the auxiliaries the deductive model requires to account for negative evidence. It would disprove too much. Recall that the cadaveric hypothesis does not itself explain all the relevant contrasts, such as why some women in the Second Division contracted childbed fever while others in that division did not, or why some women who had 'street births' on their way to the hospital contracted the fever while others did not. If the other hypotheses were rejected because they, along with *ceteris paribus* clauses, entail that there ought to be no difference in mortality between the divisions, then the model does not help us to understand why similar clauses did not lead Semmelweis to reject the cadaveric hypothesis as well.

From the point of view of Inference to the Best Explanation, we can see that there are several general and related reasons why the hypothetico-deductive model does not give a good description of the way causal hypotheses are disconfirmed by contrastive data. The most important is that the model does not account for the negative impact of explanatory failure. Semmelweis rejected hypotheses because they failed to explain contrasts, not because they were logically incompatible with them. Even on a deductive-nomological account of explanation, the failure to explain is not tantamount to a contradiction. In order to register the negative impact of these failures, the hypothetico-deductive model must place them on the Procrustean bed of logical incompatibility, which requires auxiliary statements that are not used by scientists and not usually available even if they were wanted. Second, the hypothetico-deductive model misconstrues the nature of explanatory failure, in the case of contrastive explanations. As we saw in chapter 3, to explain a contrast is not to deduce the conjunction of the fact and the negation of the foil, but to find some causal difference. The hypotheses Semmelweis rejects do not fail to explain because they do not entail the contrast between the divisions: the cadaveric hypothesis does not entail this either. They fail because they do not mark a difference between the divisions, either initially or after manipulation. Third, the model does not reflect the need, in the case of explanatory failure, to judge whether this is due to incompleteness or error. In the model, this decision becomes the one of whether we should reject the hypothesis or the auxiliaries in a case where their conjunction contradicts the evidence. This, however, is not the decision Semmelweis has to make. When he had all the mothers in both divisions deliver on their sides, and found that this did not affect the contrast in mortality, he did not have to choose between saying that the hypothesis that delivery position is a cause of fever is false and saying that the claim that everything else was equal is false. After his experiment, he knew that not everything else was equal, but this left him with the question of whether he ought to reject the delivery hypothesis or just judge it to be incomplete.

The failures of the hypothetico-deductive model to capture the force of disconfirmation through explanatory failure also clearly count against Karl Popper's account of theory testing through falsification (1959). Although he is wrong to suppose that we can give an adequate account of science without relying on some notion of positive inductive support, Popper is right to suppose that much scientific research consists in the attempt to select from among competing conjectures by disconfirming all but one of them. Popper's mistake here is to hold that disconfirmation and elimination work exclusively through refutation. As the Semmelweis example shows, scientists also reject theories as false because, while they are not refuted by the evidence, they fail to explain the salient contrasts. Moreover, if my account of the way this sort of negative evidence operates is along the right lines, this is a form of disconfirmation that Popper's account cannot be modified to capture without abandoning his central proscription on positive support, since it requires that we make a positive judgment about whether the explanatory failure is more likely to be due to incompleteness or error, a judgment that depends on inductive considerations.

The hypothetico-deductive model appears to do a better job of accounting for Semmelweis's main positive argument for his cadaveric hypothesis, that disinfection eliminated the contrast in mortality between the divisions. Suppose we take it that the cadaveric hypothesis says that infection with cadaveric matter is a necessary cause of childbed fever, that everyone who contracts the fever was so infected. In this case, the hypothesis entails that where there is no infection, there is no fever which, along with plausible auxiliaries about the influence of the disinfectant, entails that there should be no fever in the First Division after disinfection. But this analysis does not do justice to the experiment, for three reasons. First of all, the claim that cadaveric infection is strictly necessary for fever, which is needed for the deduction, is not strictly a tenable form of the cadaveric hypothesis, since Semmelweis knew of some cases of fever, such as those in the Second Division and those among street births, where there was no cadaveric infection. Similarly, given that disinfection is completely effective, this version of the hypothesis entails that there should be no cases of fever in the First Division after disinfection, which is not what Semmelweis observed. What he found was rather that the mortality in the First Division went down to the same low level (just over 1 percent) as in the Second Division. As Hempel himself observes, Semmelweis eventually went on to 'broaden' his hypothesis, by allowing that childbed fever could also be caused by 'putrid matter derived from living organisms' (1966: 6); but if this is to count as broadening the hypothesis, rather than rejecting it, the original cadaveric hypothesis cannot have been that cadaveric infection is a necessary cause of the fever.

The second reason the deductive analysis of the disinfection experiment does not do justice to it is that the analysis does not bring out the special probative force of the contrastive experiment. Even if we suppose that cadaveric infection is necessary for fever, the hypothesis does not entail the *change* in mortality, but only that there should be no fever where there is disinfection, since it does not entail that there should be fever where there is no disinfection. But it is precisely this contrast that makes the experiment persuasive. What the hypothetico-deductivist could say here, I suppose, is that the change is entailed if we use the observed prior mortality as a premise in the argument. If cadaveric infection is necessary for fever, and if there was fever and infection but then the infection is removed, it follows that the fever will disappear as well. Even this, however, leaves out an essential feature of the experiment, which was the knowledge that, apart from disinfection, all the antecedents of the diachronic fact and foil were held constant. Finally, what makes the cadaveric experiment so telling is not only that it provides evidence that is well explained by the cadaveric hypothesis, but that the evidence simultaneously disconfirms the competitors. None of the other hypotheses can explain the temporal difference, since they all appeal to factors that were unchanged in this experiment. As we have seen in our discussion of negative evidence, however, the deductive model does not account for this process of disconfirmation through explanatory failure, and so it does not account for the way the evidence makes the cadaveric hypothesis the best explanation by simultaneously strengthening it and weakening its rivals.

I conclude that Inference to the Best Explanation, linked to an account of contrastive explanation that provides an alternative to the deductive-nomological model, is an improvement over the hypothetico-deductive model in its account of the context of discovery, the determination of relevant evidence, the nature of disconfirmation and the special positive support that certain contrastive experiments provide. In particular, Inference to the Best Explanation is an improvement because it allows for evidential relevance in the absence of plausible deductive connections, since contrastive explanations need not entail what they explain. If Inference to the Best Explanation is to come out as a suitable replacement for the hypothetico-deductive model, however, it is important to see that it does not conflict with the obvious fact that scientific research is shot through with deductive inferences. To deny that all scientific explanations can be cast in deductive form is not to deny that some of them can, or that deduction often plays an essential role in those that cannot be so cast. Semmelweis certainly relied on deductive inferences, many of them elementary. For example, he needed to use deductive calculations to determine the relative frequencies of fever mortalities for the two divisions and for street births. Moreover, in many cases of causal scientific explanation, deduction is required to see whether a putative cause would explain a particular contrast. One reason for this is that an effect may be due to many causes, some of which are already known, and calculation is required to determine whether an additional putative cause would explain the residual effect. Consider, for example, the inference from the perturbation in the orbit of Uranus to the existence of Neptune. In order to determine whether Neptune would explain this perturbation, Adams and Leverrier had first to calculate the influence of the sun and known planets on Uranus, in order to work out what the perturbation was, and then had to do further calculations to determine what sort of planet and orbit would account for it (cf. Grossner 1970). As Mill points out, this 'Method of Residues' is an elaboration of the Method of Difference where the negative instance is 'not the direct result of observation and experiment, but has been arrived at by deduction' (1904: III.VIII.5). Through deduction, Adams and Leverrier determined that Neptune would explain why Uranus had a perturbed orbit rather than the one it would have had if only the sun and known planets were influencing its motion. This example also illustrates other roles for deduction, since calculation was required to solve Newton's equations even for the sole influence of the sun, and to go from the subsequent observations of Neptune to Neptune's mass and orbit. This particular inference to the best contrastive explanation would not have been possible without deduction.

Let us return now to the two challenges for Inference to the Best Explanation, that it mark an improvement over the hypothetico-deductive model, and that it tell us more than that inductive inference is often inference to the likeliest cause. I have argued that the Semmelweis case shows that Inference to the Best Explanation passes the first test. It helps to show how the account passes the second test, by illustrating some of the ways explanatory considerations guide inference and judgments of likeliness. Although Semmelweis's overriding interest was in control rather than in understanding, he focused his inquiry by asking a contrastive explanatory question. Faced with the brute fact that many women were dying of childbed fever, and the many competing explanations for this, Semmelweis did not simply consider which explanation seemed the most plausible. Instead, he followed an organized research program based on evidential contrasts. By means of a combination of conjecture, observation, and manipulation, Semmelweis tried to show that the cadaveric hypothesis is the only available hypothesis that adequately explains his central contrast in mortality between the divisions. This entire process is governed by explanatory considerations that are not simply reducible to independent judgments of likeliness. By asking why the mortality in the two divisions was different, Semmelweis was able to generate a pool of candidate hypotheses, which he then evaluated by appeal to what they could and could not explain, and Semmelweis's experimental procedure was governed by the need to find contrasts that would distinguish between them on explanatory grounds. When Semmelweis inferred the cadaveric hypothesis, it was not simply that what turned out to be the likeliest hypothesis also seemed the best explanation: Semmelweis judged that the likeliest cause of most of the cases of childbed fever in his hospital was infection by cadaveric matter *because* this was the best explanation of his evidence.

The picture of Inference to the Best Explanation that has emerged from the example of Semmelweis's research is, I think, somewhat different from the one that the slogan initially suggests in two important respects. The slogan calls to mind a fairly passive process, where we take whatever data happen to be to hand and infer an explanation, and where the central judgment we must make in this process is which of a battery of explanations of the same data would, if true, provide the loveliest explanation. But as the example shows, Inference to the Best Explanation supports a picture of research that is at once more active and realistic, where explanatory considerations guide the program of observation and experiment, as well as of conjecture. The upshot of this program is an inference to the loveliest explanation but the technique is eliminative. Through the use of judiciously chosen experiments, Semmelweis determined the loveliest explanation by a process of manipulation and elimination that left only a single explanation of the salient contrasts. In effect, Semmelweis converted the question of the loveliest explanation of non-contrastive facts into the question of the only explanation of various contrasts. Research programs that make this conversion are common in science, and it is one of the merits of Inference to the Best Explanation that it elucidates this strategy. And it is because Semmelweis successfully pursues it that we have been able to say something substantial about how explanatory considerations can be a guide to inference without getting bogged down in the daunting question of comparative loveliness where two hypotheses do both explain the same data. At the same time, this question cannot be avoided in a full assessment of Inference to the Best Explanation, since scientists are not always as fortunate as Semmelweis in finding contrasts that discriminate between all the competitors. Accordingly, I will attempt partial answers in later chapters. First, however, I will consider in the next chapter the resources of Inference to the Best Explanation to avoid some of the over-permissiveness of the hypothetico-deductive model.