Chapter 6

# MEMORY AND A POSTERIORI INFERENCE

Empiricists have always regarded observation and memory as the fundamental input for empirical knowledge; they say that further knowledge is obtained mainly by a posteriori inference. In the last chapter I discussed the sort of knowledge that can be obtained by observation. After just a few pages it became obvious that observational knowledge is very closely related to inferential knowledge; in fact, much of what we think we know by observation actually requires some kind of a posteriori inference. It will turn out that memory knowledge is basically the same as observational knowledge in this respect. A posteriori inference is therefore fundamental to most empirical knowledge. I shall be principally concerned with salient varieties of a posteriori inference in this chapter, but I begin with a brief discussion of memory and eventually conclude with an empiricist defense of the presumptions on which our everyday claims about the external world are ultimately based.

## Memory as A Source of Knowledge

Although empiricists have always regarded memory as an indispensable source of our empirical knowledge, they have generally praised memory with a slightly bad conscience. The problem with memory is that it purports to tell us something about occurrences that do not exist when the memory occurs. A recollection occurs in the present; the occurrence it refers to took place in the past. But the past is over and done with, and the same is true of the occurrences we ostensibly remember. Any correspondence between a current memory and a past occurrence cannot therefore be ascertained by direct inspection. Since we cannot infer such a correspondence by a priori reasoning, there is no inconsistency in the supposition that the world came into existence just a moment ago, complete with the recollections we now have. This possibility cannot apparently be ruled out by a posteriori reasoning, at least if that reasoning is the sort of thing empiricists have taken it to be: a matter of generalizing from experience. If we never directly experience a connection between present impression and past occurrence, we have absolutely no basis for any such generalization. By what right can we trust memory if we have no way of proving that it is ever correct?

I have just rehearsed the classic case for a skeptical view of memory. At first sight, it is a very impressive case, but reflection shows that it is highly exaggerated. Do we not sometimes have direct perceptual access to part of the past? If we can perceive things in motion, we must have this access. When someone smiles, waves a hand, or throws a ball, the action takes time, and the first part of the action is over (a past occurrence) when the last part occurs. The perception of any movement is thus attached to a temporal interval that includes past, present, and future. The past and future occurrences are past and future in a relative sense: in an occurrence with three distinguishable segments A, B, and C, the segment B is a past occurrence in relation to C but a future occurrence in relation to A. This relative futurity is a genuine kind of futurity because when A begins to occur, B has not yet taken place. When B does take place, A has taken place.

Saint Augustine, who was seriously perplexed about time, had a very different view of the present. As he explained it,

> If an instant of time be conceived which cannot be divided into the smallest particles of moments, this only is it which may be called present…. For if it be, it is divided into past and future. The present has no space.[1]

If Augustine was right here, the present is a timeless moment. But we are not conscious of a timeless moment when we make observations. The world we perceive or otherwise experience is always in motion: it is consciously changing, and we experience it as changing. The idea of a "spaceless" present is created by a process of abstraction, by thinking away the terminal elements of an experienced moment. The idea thus created is comparable to the idea of an imperceptible triangle. Our knowledge of it results from inference, not observation. The moments we observe contain relative pasts and futures as well as extended nows.

As strange and possibly puzzling as the idea of an extended present may be, we have as much reason to believe we perceive extended moments containing changing things as we have to believe that we perceive anything at all. Of course, the amount of the past included in what we can perceive is relatively little: I can perceive a smile, but I probably do not perceive the whole of a forward pass in football. I watch a quarterback move his arm forward, I see the release, and I

---

[1] Augustine (1952), p. 226.

watch the ball move down the field. I suppose it is more accurate to say that I observe a series of movements. When I observe the later movements, I am recalling the earlier ones. I know what these are like because I remember observing them.

Philosophers of an earlier time devoted a lot of attention to the phenomenology—the qualitative aspects—of memory experience.[2] Thinking about this aspect of memory is no longer fashionable in philosophy; perhaps it is now supposed that the experience of remembering may be different in different people. But the qualitative character of remembering does not seem important for epistemology anyway. What is important is the truth or probable truth of the claims people make about what they remember. People who say they vividly remember certain things are generally very confident in the truth of what they say; those who declare that their memories are dim are usually less confident; and those with memories of intermediate vividness have intermediate confidence in what they say they recall. Of course, some people are naturally more cautious or more conscientious than others; some are even more interested in calling attention to themselves or in telling a good story than in being right. The timid claims of some are therefore sometimes more trustworthy than confident claims of others.

When people claim to recall things they once experienced, their recollections can often be supported or criticized by other records of the past—by diaries, letters, photographs, films, and the like. Because of such things, we do not have to rely entirely on a person's words for our picture of what actually happened. Yet words are centrally important for many past occurrences. What Tom promised Ted or what Sally told her students on Friday could never be known in any other way. To decide whether this or that person's recollection is correct, we must in fact consider the variety of factors that are pertinent to the assessment of an observation report. Since people ostensibly remember what they saw, heard, tasted, or learned in some way, the truth of what they remember depends crucially on the truth of what they think they perceived or otherwise learned. In assessing the probable truth of memory impressions or reports, we must therefore take into account the considerations pertinent to evaluating an observation report in addition to those specifically applicable to the reliability of a subject's memory and the motives he or she may have for embellishing or even falsifying a true recollection. If the truth of a certain memory claim or the occurrence of an event a person was in a position to recall is very important, as it commonly is in a legal proceeding, we might insist on having the subject cross-examined by a competent lawyer. Not only will our existing evidence be tested by the cross-examination, but further evidence will also be obtained. As every reader of mysteries knows, the process of discovering what actually happened on this or that occasion can be extremely complicated.

In spite of the complications that I have just emphasized, most memory claims might be described as past-tense observation claims: they may differ from a typical observation claim by no more than "I see Spot run" differs from "I saw Spot run." I emphasized the variety of considerations pertinent to evaluating an observation claim in chapter five; the complications pertinent to assessing a typical memory claim may be no greater than those pertinent to its present-tensed cousin. In fact, if a person is seriously questioned about what he or she now observes, the duration of the questioning may easily convert the target of the investigation into a memory claim. What is logically special about a memory claim—what makes it deserving of separate treatment—is that the inferences properly supporting its truth or probability are essentially backward looking. A fact about the past is inferred from facts about the present.

What sort of inference is capable of providing this kind of support? Hume considered it experimental. As I explained in chapter five, the reasoning Hume called "experimental" is causal inference; it consists in inferring one fact from another by means of a causal principle obtained from experience. If we represent a certain causal principle by "As cause Bs," we can identify two associated forms of experimental inference. One infers Bs from As, or effects from causes; another infers As from Bs, or causes from effects. When Hume treats particular causal inferences in detail, as he does in his Dialogues Concerning Natural Religion, he emphasizes that the causes and effects appropriate to a given causal principle have to be identified very carefully; but his basic idea is that "all experimental reasonings are founded on the supposition that similar causes prove similar effects and that similar effects prove similar causes."[3] If Hume is right, the inferences supporting the truth of memory claims are experimental inferences of his second kind: a past cause is inferred for present effects.

The cause that is inferred by such an inference is the occurrence that is ostensibly remembered. What effects provide a proper basis for such an inference? If the emphasis is squarely on memory and not other effects of past occurrences, the relevant effect is probably the subject's memory experience or memory belief. Suppose I have the experience of ostensibly remembering (or seeming to remember) parking my car a half hour ago in section C4 of the

[2] See Russell (1948), pp. 226-232.
[3] Quoted in Kemp Smith (1948), p. 147.

parking lot outside the restaurant where I am now having lunch. If I can assume that this kind of experience is probably caused by the sort of actual occurrence it seems to represent, I can conclude that I (probably) did park my car a half hour ago in section C4 of that parking lot. In Hume's view, all the inferences that provide empirical support for the truth or probable truth of memory claims are of this causal kind. If a given claim receives additional support from a note written in a diary, the note must be viewed, if Hume is right, as an indirect effect of the occurrence it describes.

I noted in chapter five that Hume thought the causal principles used in experimental inferences resulted from experience but were not themselves inferred from any premises at all. Later empiricists generally disagreed with Hume on this last point; the consensus was, and possibly still is, that causal principles are obtained by induction, an inferential process also known as inductive generalization or enumerative induction. The skeptical view of memory that I described early in this chapter was based on the idea that inductive inferences of this kind are rationally unjustifiable. If inductive inferences can be justified, memory claims can no doubt often be justified in the way Hume thought—by causal reasoning. But there are many problems with induction. Now is a good time to see what they are.

**What is Induction?**

One of the problems about induction is how the rule should be formulated. One way of presenting this problem is to show the defects of a commonly offered formulation. The one I shall begin with is given by William Lycan (1988), and it is similar to a formulation used by Laurence BonJour. Both writers accompany their formulations with qualifying remarks, Lycan's identifying fallacious applications of the rule. His rule is this:

n% of all the observed Xs have been F.
Therefore [probably] roughly n% of all Xs are F.[4]

The fallacious applications he has in mind, which his formulation does not itself rule out, occur when the number of observed Xs is too low to represent a "sufficient" sample or when the observed Xs constitute a biased one. If your evidence class—that is, the class of observed Xs—were very small, you would normally have a very poor basis for claiming that approximately n% of all Xs have the property F.[5] Similarly, if the Xs you observe were not selected in some impartial or fair way, you would normally have a very poor basis for making a comparable claim.

In view of these fallacies, it is important to look for a rule that disallows them. Consider the following:

n% of all the observed Xs have been F.
A representative example of Xs have been observed.
Therefore, [probably] n% of all Xs are F.

This formulation would no doubt disallow generalizations from insufficient and biased samples, but to apply it, we would have to know how we are to identify a representative sample. Suppose we are told that a sample of Xs is representative of a larger reference class with respect to the frequency of having F just when the percentage of Xs having F in the evidence class is approximately the same as the percentage of Xs having F in the reference class. If this is what we are to understand by a representative sample of Xs having F, the revised inference schema would be deductively valid: the corresponding conditional statement would be analytically true. This would give us an unquestionably valid form of "inductive" inference, but we would have no way of knowing when a particular evidence class is representative in the specified sense.

In Human Knowledge: Its Scope and Limits, the only one of his many books in which he seriously discussed inductive inference, Bertrand Russell offered the following as an inductive rule:

> Given a number n of $\alpha$'s which have been found to be $\beta$s, and no $\alpha$ which has been found to be not a $\beta$, then the two statements: (a) "the next $\alpha$ will be $\beta$", (b) "all $\alpha$'s are $\beta$'s", both have a probability which increases as n increases, and approaches certainty as a limit as n approaches infinity.[6]

---

[4] Lycan (1988), p.179.
[5] The word "normally" appears in this and the following sentence for a reason that will become evident as the discussion proceeds.
[6] Russell (1948), p. 419.

To apply this rule we do not have to know whether we have a representative sample of αs and βs, but the rule will give us little help if we are interested in drawing a conclusion about the percentage of native-born Norwegians having blond hair. To draw a conclusion about a reference class that, although finite, is too large to examine as a whole, we shall need some way of estimating the size of an acceptable evidence class and of identifying an impartial way of selecting its members. Unfortunately, no general description of how these tasks may be accomplished appears to be available.

A possible reason for the dearth of general descriptions is that dramatically different sample sizes and methods of selection appear to be acceptable in different cases. Consider the way new models of automobiles are evaluated each year by Consumer Reports. Normally, just one example of a given model is examined, and the example is obtained merely by buying it from some randomly chosen dealer without disclosing the actual identity of the buyer. Although one might initially suppose that a single example is far too small to be an acceptable evidence class, reflection shows that a single example is almost certain to be representative of the model to be sold with respect to traits deemed important for the entire class. The reason for this is that automobiles are mass-produced objects subject to standard quality controls. Some manufacturers produce more reliable products than others do, but a given manufacturer is apt to produce instances of a particular model in the same way using basically the same materials. Anomalies occur, of course, but one instance can be expected to be substantially similar to any other instance of the same model, particularly if the dealer selling it has no reason to suppose that the buyer will use a particular instance in a way that will compromise future sales.

The acceptability of the sample size and the method of selecting instances in this last case obviously depend on background information about automobiles and the way they are produced. Here the acceptability of one inductive inference appears to depend on the acceptability of others. The question therefore arises, "Is there is a basic rule for rationally compelling enumerative induction whose application does not require background knowledge of this kind?" As far as I know, the answer is no. Hume, assuming no relevant prior knowledge of the objects of an empirical generalization, argued that an evidence class, no matter how selected and how extensive it may be, provides no rational basis for the conclusion that the objects of the reference class, many of whose members may exist in the distant past or the remote future, are at all similar (in the relevant ways) to those already examined. We naturally expect them to be similar—"we expect the future to be [relevantly] like the past", he said—but this expectation has no basis in reason or any operation of the understanding. It is purely instinctive. There is no inconsistency supposing that the future will be unlike the past in relevant respects; and any a posteriori reason that could be offered to dispute this would be based on the same supposition and thus beg the question at issue.[7]

### Induction: Arguments Pro and Con

Laurence BonJour recently countered Hume's criticism with an a priori argument, one featuring a form of inference that Hume did not consider. BonJour's a priori argument, which is of course a defense of enumerative induction, is noteworthy for two basic reasons. It includes novel qualifications to the inductive rule designed to avoid objections raised only in recent times, and it relies on an additional form of inference that is now fashionable with philosophers and deserving of critical attention.

BonJour's argument applies to what can only be called a very incomplete formulation of an inductive rule. Initially, he identifies the sort of situation in which an inductive inference could (as he sees it) be reasonably made. The situation would involve "a large number of observed instances" of something A, a fraction m/n of which have "some logically independent observable property" B. The locations and times of observation, the identity of the observers, the conditions of observation, and any further pertinent background circumstances must be varied "to a substantial degree" and there must be no relevant background information available concerning either the incidence of Bs in the class of As or the connection, if any, between being A and being B."[8] If these conditions are met and the observed proportion of As that are Bs "converges over time to the fraction m/n and thereafter remains at least approximately constant as significant numbers of new observations come in," then the conclusion of the argument is likely to be true (p. 207).

As I noted, BonJour attaches some important qualifications to the inference he describes here, but his a priori argument for its acceptability—for the fact that its conclusion is probably true when its premises are true—is that the truth of the conclusion provides the best explanation for the data that the premises describe. The crucial data here pertain to the "convergence and

[7] See Hume, *Enquiry,* Sect. IV.
[8] BonJour, pp. 188f.

constancy of the observed proportion," and although it is possible, he says, that this proportion is a matter of chance, it is highly likely that the observed proportion is an accurate reflection of an objective regularity (p. 209). Such a regularity cannot be a mere constant conjunction, as Hume thought; to provide the required explanation it must be a "metaphysically robust" regularity, involving a necessary connection (p. 215) or a "substantial propensity to persist into the future" (p. 214).

The qualifications BonJour adds to his account apply to two kinds of counterexamples independently discovered by Bertrand Russell and Nelson Goodman in the late 1940s. Both philosophers observed that any objects chosen as the basis for an inductive generalization possess some features that support objectionable generalizations—generalizations that are either patently false or incompatible with other generalizations that are equally well supported by the available evidence.

Russell's examples showed his usual wit. One was based on the well-known belief that Immanuel Kant had never been more than ten miles from his hometown of Königsberg.[9] If Kant had been interested in drawing inductive conclusions about sheep, one property that he might have observed in every sheep he examined is that of being within ten miles of Königsberg. To get as large a sample as possible, he could have devoted years to the task of observing sheep and, to make his selection as unbiased as possible, he might have observed them in fields, in barns, on houses, and possibly even in ponds. A generalization supported by his observations would have been the patently false "All sheep are within ten miles of Königsberg." Other properties possessed by every sheep he might have observed are being observed by Immanuel Kant, being observed by someone, living in Germany, living in Europe, and being outside of Italy. Obviously, this list could be extended indefinitely.

Goodman's examples featured contrived predicates such as "grue," the latter applying to an object, Goodman stipulated, just when it is either green and examined before a distantly future time t or blue and not so examined. Goodman argued that if we are examining emeralds for color and find that they are invariably green, we can use the inductive principle to draw two incompatible conclusions, neither of which is better supported than the other. They are "All emeralds are green" and "All emeralds are grue." These conclusions are incompatible because they disagree about the color of emeralds not examined before t. According to one, they are green; according to the other, they are blue. These conclusions are equally well supported because every emerald we examine will inevitably be examined before t and thus, if it is green, count as grue. Since the time t may be placed arbitrarily far in the future, we cannot avoid the difficulty by waiting to see how emeralds look when t arrives. We are, in fact, faced with a general problem arising from the inductive principle itself. It permits us to draw incompatible conclusions from the same body of data. This defect is illustrated by the hypothesis featuring the word "grue," but it is common to countless other hypotheses. To avoid it, the inductive principle needs significant qualification.

As you might expect, one of BonJour's qualifications requires that the predicates used in an acceptable inductive inference do not include the contrived sort exemplified by "grue" (p. 189, note 2). Unfortunately, this qualification is not actually effective in avoiding the problem Goodman raised. The choice of predicates is in fact irrelevant to the issue. Instead of using the predicate "grue," Goodman could raise his problem simply by speaking of things that are either green before the time t or blue afterwards. If every emerald we examine is green, every emerald we examine is either green before t or blue after t. Because of this, the emeralds support the hypothesis H2, that all emeralds are either green before t or blue afterwards, just as strongly as they support the hypothesis H1, that all emeralds are green. But H2 conflicts with H1 in regard to the color emeralds have after t. This problem is evidently not avoided by the inductive rule that BonJour attempts to justify. The problem does not arise, incidentally, from the disjunctive character of one of the hypotheses. It can arise just as easily from curve fitting problems.[10]

When Goodman introduced predicates such as "grue," his aim was to call attention to what he called a new problem of induction. This new problem was essentially the same as the one Russell raised: the familiar inductive rule needs serious qualification if it is to be acceptable. Russell did not suggest a qualification for the standard rule; Goodman did. Oddly enough, the qualification Goodman offered bears an interesting similarity to the principal qualification BonJour provided. Somewhat like BonJour, Goodman ruled out troublesome hypotheses—the one's containing words with the same extension as predicates such as "grue" [11]—on the ground that they are not "lawlike."[12] He distinguished lawlike from non-lawlike hypothesis by reference

---

[9] Russell used this last property to illustrate the "shaky" character of induction by simple enumeration in Russell (1951), p. 126.
[10] See Grunstra (1969), pp. 102-106.
[11] Predicates P1 and P2 have the same extension just when they apply to the same objects.
[12] Goodman (1965), p. 73.

to a property that he called "entrenchment," which BonJour did not come close to mentioning, but he did imply that the hypotheses strongly confirmed by their instances could be regarded as laws, or statements of laws. This recalls BonJour's claim that regularities inferable by enumerative induction must be "metaphysically robust" regularities, involving necessary connections or "substantial propensities to persist into the future."

### Induction and Laws

The idea that the generalizations reasonably inferred from data samples are or must be laws is so implausible that it is hard to take seriously. People who conduct public opinion polls draw general conclusions from their samples, but they rarely if ever suppose that their conclusions hold true eternally or even far into the future. The same is true of conclusions about the effects of advertisements, the fear of epidemics, or any of the thousands of topics that are investigated by statistical methods every year. There is nothing "lawlike" about the conclusion that the U.S. President's approval rating among voters is ten per cent less today than it was two months ago, even though this conclusion was inferred from samples taken all over the country.

Some of the generalizations inferred from experimental data might, I suppose, be considered "laws," although the very idea of a scientific law is less widely accepted these days than it used to be.[13] But there is no agreement among statisticians that inferred laws are generally more secure than short-term generalizations about public opinion.[14] If this is right, then if enumerative induction deserves to be regarded as an acceptable form of a posteriori inference, BonJour's a priori justification at best applies only to a limited class of these inferences—and not to a favored class whose members are used with greater confidence than the others. The qualifications he places on the kinds he defends do rule out some of the counterexamples Russell constructed, but they do not succeed against Goodman's counterexamples, which do not really depend on special predicates, nor do they succeed against the full range of counterexamples Russell had in mind, which are essentially the same as Goodman's.[15]

Like BonJour, Goodman wanted to disallow the "bent" hypotheses he discussed as well-confirmed examples of inductive conclusions, and he did so, I said, by claiming that they are not lawlike and so not confirmable by their instances. But Goodman's solution to his new riddle of induction is arguably too restrictive even for the case of scientific "laws". Citing specific scientific theories, Rosenkrantz (1981)
persuasively argued that scientific advances often result in hypotheses that are more "bent" than the ones they supersede. The grue hypothesis, he said, in fact "belongs to a class of hypotheses that are not only scientifically quite respectable but are the very ones whose introduction so often marks the breakthroughs we are wont to label 'scientific revolutions'" (p. 7.1, 4). The price of adopting Goodman's "entrenchment" solution to the new riddle, Rosenkrantz contends, is much too great to tolerate.

An enormous literature has grown up around Goodman's new riddle and his proposed solution to it, and Rosenkrantz's criticism, as impressive as I find it, is no doubt not the last word on the matter. Specialists in the history and methodology of science can speak to it far more effectively than I can. But BonJour's a priori defense of induction has another feature that raises important issues of a different kind. It is based on what is now known as an inference to the best explanation, a form of inference that is widely regarded as a posteriori rather than a priori. I will discuss the logical structure of this kind of inference a little later; right now I want to say something about BonJour's belief that the inference he employs in defending a schematic example of enumerative induction is a priori.

Judging by the steps he takes in constructing his argument, I think it is fair to say that BonJour's belief in this matter rests on two assumptions, which he thinks he knows to be true a priori:[16]

1. The best explanation (meaning "best explanatory account") that can be given for a body of data is most likely to be true.

2. The best explanation that can be given for the truth of a standard inductive premise is the straight inductive explanation, namely that the observed proportion m/n reflects (within a reasonable degree of approximation) a corresponding objective regularity in the world.

---

[13] See van Fraassen (1989), Part 1, pp. 15-128.
[14] See Phillips (1974), chapter 6, for an elementary discussion of probability densities.
[15] See Russell (1948), p. 422.
[16] See BonJour's principle (I-2) in BonJour (1998), p. 212.

In formulating these two assumptions I am ignoring some claims that BonJour makes in the course of his argument but that are actually not needed for it.[17] The "best explanation" that he identifies in (2) is clearly the best that can be given in his opinion; and the truth of (1) is something he thinks he can simply see to be true. The assumption I now want to say something about is (2).

BonJour thinks what he calls "the straight inductive explanation" (or SIE) is the best one for two reasons. First, he thinks the connection between A and B observed in the evidence class must be explainable by some law. And second, he thinks that any genuine law consistent with the evidence but requiring a divergence from the observed ratio m/n in a way that would falsify SIE would not really be possible. A genuine law requiring a divergence from m/n could be owing only to a further characteristic C, he says, one that affects the facts of observation itself, and this runs afoul of one of the qualifications he mentioned in describing acceptable induction in the first place.

These two reasons are idiosyncratic and certainly not convincing. As for the first, I can think of no tenable basis for supposing that a regularity observed to hold during some finite interval, however long, can be explainable only by a law. A more extensive regularity, one without temporal limits, will certainly do as well. An actual law is not needed, I should say, because explanation, pragmatic considerations aside, is prediction after the fact, and anything predictable by a law is equally predicable by a temporally unrestricted regularity: the modal character of a law, its supposed necessity, has no distinctive observable consequences. As for the second reason, an appropriate regularity between A and B can certainly be such that m/n of As are B in one spatial or temporal region but j/n of As are B elsewhere (j being significantly larger or smaller than m). The variation can simply be a matter of the way A is related to B; another characteristic is not needed to account for the divergence from m/n. The sort of "bent" hypotheses (or supposed laws) that Rosenkrantz cites in criticizing Goodman provide actual examples of such lawfully predicable divergences.[18]

If I am right about these last points, BonJour's attempted priori justification of enumerative induction does not succeed; it does not even overcome the arguments casting doubt on the idea that enumerative induction does not deserve to be considered an acceptable form of inference. But there is a further matter to be discussed, the acceptability of the form of inference that BonJour relied on in his attempted justification--namely, Inference to the Best Explanation or IBE. This form of inference is now widely accepted; in fact, some well-known writers—for example, William F. Lycan—regard IBE as the basic form of a posteriori inference. According to Lycan, enumerative induction can be reconstructed as a special case of IBE.[19]

## Inference to the Best Explanation

To evaluate this form of inference, one must understand its logical structure. Lycan describes this structure as follows: [20]

1. F1,…,Fn are facts.
2. Hypothesis H, if true, would explain F1,…,Fn.
3. No available competing hypothesis would, if true, explain the Fi as well as H does.
4. Therefore, [probably] H is true.

There is no doubt that we often reason according to this pattern in everyday life, but it is actually very doubtful whether the practice is as commendable as Lycan supposes.[21] The third premise should raise immediate doubts. What count as competing hypotheses in a particular case? If we are to apply the method, we must be able to survey all the available "competing"

[17] One assertion that does no work in his argument is that "it is highly likely that there is some explanation (other than mere coincidence or chance) for the convergence and constancy of the observed proportion…." (p. 208). This is obviously not needed if he can simply identify the best explanation.

[18] Rosenkrantz (1981) describes these hypotheses as positing "theoretically well-founded deviations from an overriding 'straight' hypothesis at extreme ranges of the relevant variables." His examples are taken from both special and general relativity. See Rosenkrantz, ch. 7, sect. 1. See also Rosenkrantz's lucid paper (1982), which deserves to be considered a classic on the philosophy of induction.

[19] Lycan (1988), 178-188. A more extensive discussion is given in Lipton (2004), but Lycan's description is adequate for the task at hand. Bas C. van Fraassen and others discuss critically the second edition of Lipton's book (2004) in van Fraassen (2006). See below, footnote 21.

[20] *Ibid*. p. 129. I alter Lycan's description in trivial ways compatible with his intent. He uses "explain" where I use "would explain if true," but he says he is using "explain" in the "nonsuccess" sense equivalent to "would explain if true."

[21] van Fraassen (1989) criticizes this form of inference in some detail. His objections are in general agreement with the objections I develop here. See also his contribution to the review symposium on the second edition of Lipton's *Inference to the Best Explanation* in van Fraassen (2006), pp. 344-352.

hypotheses. If a pertinent hypothesis is overlooked, we cannot be sure we have found the best explanatory account. The class of competing hypotheses must therefore be limited to those we can think of; they cannot comprise a class of ideal alternatives. But why should a reasonable philosopher suppose that the right explanation for the Fi is generally supplied by one of the hypotheses some actual person can think of? It is obvious that most of the facts we can explain today by the quantum theory or the theory of relativity could not be rightly explained by any hypothesis that Plato or Aristotle could even conceive of. Are we to suppose that we are bound to be in a better position to explain an arbitrary occurrence than they were? If a phenomenon is similar to others that we have successfully explained by accepted principles, we can approach it with a fund of knowledge that may assist us in identifying the likely explanatory factors. But if we lack this knowledge, the account that seems best to us might be wide of the mark and certainly not "probably true."

As it happens, real-life inferences to the best explanation are commonly fanciful and irresponsible. People not trained to weigh evidence confidently offer explanations in cases where they lack the information to provide any reasonable explanation at all.[22] A skeptical attitude does not seem to be natural to ordinary human beings. When a fanciful explanation is offered for some fact, judicious observers are apt to reject it immediately *even if no alternative explanation is available*; they do so because they doubt the proffered account is actually true. If they do accept an explanatory account, it is only because it is already significantly credible; in choosing it, they are generally convinced that it rather than some other acceptable principle is applicable to the facts in question. If no generally accepted principle seems to apply, they may speculate about a possible explanation, but they never, if they are judicious, actually accept an explanatory hypothesis as "probably true" (at least they ought not to do so) if they merely regard it as preferable to the other explanatory accounts that they can think of. They might regard it as providing a possible explanation that deserves to be kept in mind and tested further, but they would not accept it as "probably true" if it had nothing else in its favor.

About the only time a hypothesis is regarded as strongly supported by its evident success in explaining certain observed facts is when that hypothesis is antecedently probable, that is, already acceptable in a significant degree, or the facts are antecedently improbable, not predictable by other accepted principles. An example of a hypothesis supported this way was Einstein's general theory of relativity; photographs of fixed stars taken during a solar eclipse supplied the supporting facts.[23] Before the photos were taken, the theory was not regarded as sufficiently probable to be accepted, but the disjunction "Either Einstein's theory is true or Newton's theory of light and gravitation is true" was regarded as highly probable, and not entirely owing to the probability of the Newtonian disjunct. The facts were antecedently improbable, because the fixed stars had never been observed in the precise arrangement predictable by Einstein's theory and verifiable by the photographs. Unlike the hypothesis supported in this example, one with a low antecedent probability generally receives only weak support from the data predictable by means of it. To obtain nontrivial support for such a hypothesis, persistent testing will normally be needed. Those who favor the hypothetico-deductive method emphasize the importance of such testing, but it is not even suggested by Lycan's description of Inference to the Best Explanation.

The HD or hypothetico-deductive method is a kind of precursor to IBE.[24] According to it, a hypothesis is tested by deducing consequences from it in conjunction with auxiliary assumptions that are considered true or approximately so. If the consequences are verified, the hypothesis is confirmed in a positive degree; if they are refuted, the hypothesis is amended and tested again, or simply rejected. This kind of testing is supposed to be persistent, for a hypothesis is considered acceptable only if it stands up to a considerable amount of testing. In this respect, its confirmation is similar to what a hypothesis is thought to require from enumerative induction. Understood as I have described it, the HD method seems a bit simple-minded, because the testable consequences of one hypothesis can always be inferred from another one. (This is trivially true, because an investigator can create a new hypothesis from an old one by adding on some qualification, humdrum or exotic.)[25] For this reason, the HD method is naturally modified along the lines of Inference to the Best Explanation. The aim now is to choose the best hypothesis from a family of alternatives. Testing is by prediction, as before, but now alternatives covering the same data must compete. The best hypothesis should ideally be the simplest, the most testable, and one that fits in best with background knowledge.

[22] Evidence for this is given in Tversky and Kahneman (1983).
[23] See Carnap (1956), pp. 158ff.
[24] See Aune (1970), pp. 167-182, and Earman (1983).
[25] The HD method does not require that each alternative hypothesis must be antecedently probable in a significant degree. A requirement of this kind is appropriate to the view of confirmation I discuss in the next section.

As I see it, the modified HD method, like Lycan's description of IBE, does not accord with reasonable scientific methodology. For one thing, it possesses one of the basic defects of IBE: the family of alternatives it features are just the alternatives someone can think of, and there is no a priori basis for thinking that one of these alternatives is apt to be true. For this, additional knowledge is needed. For another thing, the probability of the conclusion we can expect to infer from observed data is bound to admit of degrees: some conclusions are weakly supported at best; others are supported more strongly. The antecedent probability (or plausibility) of the other hypotheses should also be taken into account, and the same is true of the supposed facts that these hypotheses are supposed to predict or explain. Clearly, some hypotheses are more far-fetched than others, and the existence of antecedently unlikely facts predicted by a hypothesis will support it far more strongly than will those that are likely to occur anyway. A conception of experimental inference that does not accommodate these probabilistic considerations cannot be deemed satisfactory. Inferences conforming to the HD method or Lycan's pattern are generally dubious, I should say, because they ignore too much that is pertinent to the support of an acceptable hypothesis.[26]

**Inferences Based on Bayes' Theorem**

There is an alternative form of inference that does not possess the limitations I have just mentioned. It is based on the use of a simple theorem of probability theory, one known as Bayes' theorem. Ordinary people and even most scientists rarely employ this theorem in routine inferences, but then they rarely employ formal logic either. Formal logic and probability theory are indispensable when informal inferences need to be evaluated for rational acceptability—for validity or cogency. Since experimental inferences have conclusions that are more or less probable, the evaluative principles particularly appropriate to them should include the principles of probability. Bayes' theorem is a very important principle of this kind.

To understand how Bayes' theorem can be used in the evaluation and logical reconstruction of experimental inferences, we have to understand something about the principles of probability and how they can be applied to the task at hand. Studying probability theory can take you quickly into some serious mathematics, but the inferences I intend to describe here can be understood with only a minimal exposure to mathematical symbolism. I shall say just enough about the principles of probability to make an elementary use of Bayes' theorem understandable. You can understand me if you can recall the elementary parts of your high school algebra.

Formally speaking, the principles of probability comprise a remarkably simple mathematical system commonly known as the probability calculus. An important feature of this calculus (understood as a formal system) is that it can be interpreted in many different ways.[27] On one standard interpretation it applies to physical outcomes (changes in the world); on the one I shall use, it applies to statements or assertions. Applied this last way, the calculus concerns what are sometimes called epistemic or evidential probabilities.[28] As I shall understand them, these probabilities are degrees of certainty and evidential support.

The simplest probability statements of the kind in question are categorical in form; an example is "$P(p) = a$," which may be read "the probability of p equals a." The values assigned to these statements—for instance, the value represented here by "a"—are taken from the real numbers between 0 and 1 inclusive. 1 is the maximum value, indicating certain truth; 0 is the minimum value, indicating certain falsity. Since "$p \vee {\sim}p$" is certainly true and "$p \wedge {\sim}p$" is certainly false, $P(p \vee \sim p) = 1$ and $P(p \wedge {\sim}p) = 0$. The probability value of statements that are neither certainly true nor certainly false are represented by real numbers between 1 and 0; the value of statements closer to 1 are progressively more certain than those whose values approach zero and are progressively less certain than those closer to one. If we believe that a statement's degree of certainty is fairly close to 1, we might assign it a probability value of 0.9. If we think it is very uncertain, we might assign it a value of 0.2, which is tantamount to assigning its negation a value of 0.8.

These last remarks can be expanded to reassure readers not used to thinking of numerical degrees of certainty and support. Since certain truth is equivalent to a probability of 1 and certain falsity is equivalent to a probability of 0, a probability of 0.5 is equivalent to probabilistic indifference, where a statement is no more likely to be true than its negation. A probability of 0.75 is then intermediate between such indifference and certainty, so it amounts to "fairly probable" in

---

[26] Another consideration, emphasized by van Fraassen, is that any inductive principle that yields conclusions incompatible with those obtained by the probability calculus (on the same evidence) possess a kind of incoherence. He supports this consideration by a so-called Dutch book argument, which I do not discuss in this book. See van Fraassen (1989), ch. 7. For further discussion of such an argument see Skyrms (1986), ch. 6.
[27] See Skyrms (1986).
[28] *Ibid*, p. 15.

everyday terms. Probabilities over 0.9 therefore count as "high." The fact that the probability assignments resulting from informal inferences to the best explanation are commonly thought to be no more precise than "slightly probable," "quite probable," and "highly probable" suggests that numerical assignments need not, in practice, be exact either. In most cases one can think of a numerical assignment as an approximation, representing a value in the neighborhood of what the number strictly represents.

Bayes' theorem provides a principle for calculating what are known as conditional probabilities. The formula "P(q/p) = a", as I shall interpret it, may be read "the probability of q on the assumption p = a" or, more simply, "the probability of q on p = a"; it expresses the degree to which q is evidentially supported on the assumption that p is true.[29] As in the case of ordinary deduction, the evidential support represented by conditional probability is hypothetical, because a false premise does not unconditionally support a conclusion that it entails. If Rover is a dog, the statement "Rover is a cat" entails "Something is a cat" but it does not succeed in showing that the latter is true. When we say that q is evidentially supported to the degree x by the hypothesis that p, we mean that the truth of that hypothesis would provide x degrees of support for q: it would raise p's probability by that amount. (Other evidence that counts against q could, of course, undermine this degree of support if the latter is limited.)

Although, according to chapter one, evidence need not be propositional, a statement can always describe its nature or character. We can therefore use the term "P(h/e)" to denote the probability of h on the evidence e. If e entails h, the probability of h on e is maximal, as great as evidential support can be. Maximal support is represented by "1", the integer that also represents certain truth. As you would expect, maximal disconfirmation is represented by "0": if e entails ~h, P(h/e) = 0. Although conditional probability statements are ideally suited to express the degree to which a statement of evidence would, if true, support some hypothesis, they serve the more general purpose of expressing the degree to which one statement with a given probability value hypothetically supports another statement with a given value. Bayes' theorem, as I said, provides a general principle for ascertaining such a degree of support.

A simple form of Bayes' theorem can be stated as follows:

(SBT) If $P(e) \neq 0$, then $P(h/e) = P(h) \times P(e/h) \ / \ P(e)$.

This statement is much less complicated that it might initially appear. Evidence statements are almost always contingent statements, not certain falsities, so their probability is almost always positive. Thus the significant core of the theorem is the equality:

$$P(h/e) = \frac{P(h) \times P(e/h)}{P(e)}$$

The left side of the equation can be taken to denote the probability of some hypothesis on the evidence e; the right side gives the formula by means of which this probability can be calculated, namely:

$$\frac{P(h) \times P(e/h)}{P(e)}$$

This last formula is not only very simple, but it encapsulates the commendable aspects of the reasoning in the hypothetico-deductive method and in inferences to the best explanation without including their defects. Unlike these other methods, it is sensitive to three things whose importance for hypothesis testing I have emphasized: the acceptability or antecedent probability of the hypothesis being tested (represented by "P(h)"), the antecedent probability of a predicted outcome (represented by "P(e)"), and the degree to which the hypothesis hypothetically supports that outcome (represented by "P(e/h)").

Suppose that the hypothesis h (into which we can incorporate pertinent background information) predicts e with a positive degree of certainty n. If we contemplate the fraction by which the degree to which e hypothetically supports h can be calculated—namely,

$$\frac{P(h) \times P(e/h)}{P(e)}$$

[29] The formula "P(q/p)" is often glossed as "the probability of q given p," but the dangling participle in this locution is no clearer, in my opinion, than a dangler is in most other cases. I therefore prefer to avoid it here.

—we can see that the more improbable e is, that is, the lower the value of P(e), the larger the value of the fraction will be and, therefore, the greater the value of P(h/e) and the more strongly e will hypothetically support h. We can also see that the higher the credibility (or antecedent probability) of the hypothesis h, the larger the numerator of the fraction will be and therefore the larger that fraction will be. This means that (other things being equal) the greater the probability of the hypothesis, the more strongly it is supported by the occurrence of what it predicts. These facts accord beautifully with informal principles of a posteriori inference: implausible, ad hoc hypotheses that accord with observable data are not (generally speaking) strongly supported by that data; and the occurrence of antecedently unlikely data strongly supports hypotheses that predict it.

I mentioned earlier that if p entails q—that is, if it is certain that p ⊃ q—the conditional probability of q on p = 1. This principle holds for all p and q. As a result of this, if a hypothesis h entails e, the negation of e entails the negation of h—that is to say, ~e entails ~h. But if ~e entails ~h, P(~h/~e) =1. We therefore have a mathematically sound basis for the principle that the nonoccurrence of evidence predicted by a hypothesis effectively refutes that hypothesis, rendering its negation conditionally certain.

What about the problem of alternative hypotheses that I mentioned in the last section? When I discussed the hypothetico-deductive method, I observed that the occurrence of predicted data cannot accord significant support to a hypothesis by itself because data that accords with one hypothesis always accords with other hypotheses, thus adding no support to any hypothesis in particular. Does this observation undermine the usefulness of the version of Bayes' theorem we are considering? The answer is no. If, in computing the value of P(h/e), we assign a high antecedent probability to the hypothesis h, we single it out as a special hypothesis that (other things being equal) will receive significant support from data whose occurrence it allows us to predict. And if, on the other hand, we assign a low antecedent probability to e, the basis for our assignment can only be other hypotheses that we are tacitly taking account of; we are in effect assuming that e has, on the average, a low probability given the totality of hypotheses bearing upon its value.

### Ascertaining Prior Probabilities

The answer I gave in the last paragraph is bound to raise a more fundamental question: "How are we to ascertain the antecedent probabilities, the ones such as P(h), P(e/h), and P(e), that are needed to apply Bayes' theorem?" Actually, I have already given part of the appropriate answer. As far as P(e/h) is concerned, the appropriate value can often be obtained by ordinary deduction. If h is a compound formula containing not only a hypothesis but the background assumptions needed to support a prediction e, h can be expected to entail e, so that P(e/h) = 1. To make matters more perspicuous, we might identify these background assumptions explicitly, using a term such as "P(e/h ∧ a)" (meaning "the probability of e on h and a) rather than the "P(e/h)."[30] As in the simpler case of "P(e/h)," we can often calculate the value of a term such as "P(e/h ∧ a)" by ordinary deduction. If the conjunction of h and a entails e, P(e/h ∧ a) = 1; if that conjunction implies that e has a lesser value j, P(e/ h ∧ a) = j as well.

The antecedent probability of the categorical h's or e's required to apply Bayes' theorem can usually be calculated from pre-existing information concerning those assertions. Often the calculation for a hypothesis h can be made by a rule of conditioning (as it is called) applied to prior applications of Bayes' theorem. If, having used Bayes' theorem to calculate a value, say n, for the conditional probability of an assertion h on evidence e*—that is, for P(h/e*)—we may proceed to assign the value n to h itself—that is, to P(h)—if we learn that e* is in fact true. When we do this, we are updating the value of h. We needed an antecedent or "prior" probability for h in order to calculate its conditional value on the evidence of a predicted e*, but when we learn that "e*" is actually true we can give h a new "posterior" probability. This posterior probability can become a prior probability for further applications of Bayes' theorem; the designations "prior probability" and "posterior probability" are in fact applicable to an assertion only in relation to a Bayesian calculation and a verified prediction. Bayes' theorem is a powerful investigative tool because it can be applied again and again to a single hypothesis, updating its probability value as evidence accumulates. Thus, the value we assign to P(h) in an application of Bayes' theorem may often be computed by a prior application of that theorem to some prior, ascertained evidence e*.

[30] Calculating the value of this more complicated probability requires a slightly more complicated Bayesian rule, specifically:

$$P(h/e \wedge a) = \frac{P(h/a) \times P(e/h \wedge a)}{P(e/a)}, \ if\ P(e/a) \neq 0.$$

As for the probability of an evidence statement e in an application of Bayes' theorem, we can often calculate this by estimating its value in relation to background hypotheses. If j and k are incompatible hypotheses one of which is bound to be
true, we can estimate the value of P(e) by ascertaining its probability on both of these hypotheses and qualifying each conditional probability by factoring in the antecedent probability of each hypothesis. The rule to apply here, expressed symbolically, is "if j and k are jointly exhaustive, mutually incompatible hypotheses, then $P(e) = P(j) \times P(e/j) + P(k) \times P(e/k)$."[31] Here is a simple example of how the rule is applied. Suppose we have the following information about one member of a pair of dice: it is either fair or slightly biased in favor of heads, but much more likely to be fair than biased. We also know that the probability of getting heads if it is biased is 0.7 and that the probability of its being biased is 0.2. Consistency then requires us to assume that the probability of getting heads if it is fair is 0.5 and the probability that it is fair is 0.8. What, we want to know, is P(e), the probability of getting heads on a single throw of this die? The answer is $P(e) = P(f) \times P(e/f) + P(b) \times P(e/b) = 0.8 \times 0.5 + 0.2 \times 0.7 = 0.4 + 0.14 = 0.54$.

These strategies for computing values for the probabilities needed to apply Bayes' theorem have an obvious drawback from a philosophical point of view. They show us how to assign a probability value to a statement only if we already know other relevant probability values. These strategies do not therefore tell us how a basic probability value is rightly determined. Yet without basic probability values, we cannot assign a value for any probability other than a so-called likelihood—that is, a statement giving the conditional probability of an outcome on some hypothesis and background assumptions. (I have said that this kind of probability can often be determined by ordinary deduction.) How can we possibly ascertain basic probability values?

According to an influential school of statisticians known as subjective Bayesians, the basic probabilities needed for experimental inference can simply be assumed, because they do not have to be well founded or accurate in some sense. Experimental inference based on Bayes' theorem is, they say, self-correcting. If we begin with prior probabilities that are not extreme (close to zero or one) and continue to update our probability values by the rule of conditioning, the effect of our initial prior probabilities will become progressively smaller as we proceed: two people starting out with different prior probabilities and updating their probability values by successive conditioning involving the same evidential input will eventually agree on the probabilities they ascribe to relevant hypotheses. This claim, which can be demonstrated mathematically, shows that inferred probabilities can be acceptable without being based on objectively correct priors.[32]

The mathematical fact that people who update their prior probability functions by persistent conditioning on the same evidential data will eventually agree on the probability values they assign to resultant hypotheses does not really dispose of the philosophical problem at issue here. In actual cases in which the resultant agreement is approximated, there is a great deal of presupposed agreement on the admissible evidence, on the alternative hypotheses to be considered, and on such things as the probabilistic independence of occurrences pertinent to their calculations.[33] It is possible to seek experimental support for what is thus presupposed; but to obtain this support, further inferences of a probabilistic sort will have to be made, and these inferences will require further assumptions about prior probabilities and evidential data. If people with different priors disagree on any of these matters, the probabilities they eventually assign to the relevant hypotheses are not likely to be the same.

If we are to use Bayes' theorem as a basic rule of experimental inference, we must therefore find some way of justifying basic probability statements. If a statement is analytically true, it is of course certain and has a probability of 1; and if a statement p implies a statement q, the probability of q on p is also 1. Similarly, statements that are analytically false have a zero probability, and if p is inconsistent with q, the probability of q on p is also 0. In other cases, probability theory applied to statements cannot itself assign a value to any categorical statement.

It is important to realize that the limitation I have just mentioned also holds true for ordinary deductive logic: only logically true and logically false statements are given a definite value by logic itself. The value of contingent statements must be determined empirically. As regards these statements, logic can tell us only what is true, or false, if something else is true, or false. Generally speaking, the point in knowing that Q is a deductively valid consequence of a premise P is that we should be inconsistent if we accept both P and ~Q. If these propositions concern matters of fact, the choice between them is not a logical one. If we accept P, we must not accept ~Q; if we accept ~Q, we must not accept P.

[31] The antecedent here is "$(j \vee k) \wedge \sim(j \wedge k)$."
[32] See Phillips (1973), Rosenkrantz (1981), Skyrms (1986), and Howson and Urbach (1989).
[33] I show this in the Appendix to Aune (1991).

According to a subjectivist interpretation, the probability calculus places consistency conditions on statements expressing degrees of belief or confidence in propositions. As Frank Ramsey put it is his pioneering essay, "the laws of probability are laws of consistency, an extension to partial beliefs of formal logic, the logic of consistency."[34] This way of looking at the probability calculus can be illustrated by a fumbling attempt to apply Bayes' theorem. Since an antecedently unlikely occurrence strongly supports a hypothesis from which it is predictable, and since antecedently probable hypotheses are, other things being equal, more strongly supported by the predictions they warrant than antecedently improbable hypotheses, it might occur to a person beginning the study of probability that a favorite hypothesis h (to which he assigns a moderately high probability of 0.8) would be very strongly supported by a testable consequence e with a low antecedent probability of 0.4. A simple computation shows, however, that this arrangement of probabilities is inconsistent. If e is deducible from h (so that "$h \supset e$" is certain), $P(e/h) = 1$. Given this value of the likelihood $P(e/h)$, one can infer from Bayes' theorem that $P(h/e) = P(h)/P(e)$, which = 2 in this case. But this is an impossible result, since no probability can be greater than 1. Reflection shows that if p entails q, the probability of q cannot be less than the probability of p—and this fact was not appreciated in the case I have described.

**Basic Prior Probabilities**

The fact that epistemic probabilities are constrained by analytic certainties is enough to show that a purely subjectivist interpretation does not accord with the approach I have taken here. It is not a matter of subjective belief that Q is analytically true or analytically false; it is also not a matter of subjective belief that a contingent R or S is known to be true. Contingent matters cannot be known to be true for certain, so they do not deserve a probability of 1; but anything that is known to be true in a looser sense deserves a very high probability--less than 1 but reasonably close to it. Since so called likelihoods—that is, assertions of the form "$P(e/h \wedge a) = n$"— can usually be given a probability value on the basis of deduction, the remaining probability statements that need some extra-logical justification are those assigning values to basic prior probabilities, those not inferred from other probabilities. The question is, "How can these basic probability statements possibly be justified?"

An answer to this question can be located by reflecting on a basic epistemic principle laid down by that critic of empiricism, R. M. Chisholm. The principle is:

CP: If S accepts h and if h is not disconfirmed by S's total evidence, then h is probable for S.[35]

The conception of probability involved in this principle is not the conception I have been concerned with in this chapter; it is an idiosyncratic conception that Chisholm expresses by the adjective "internally probable." According to him, a proposition is internally probable for a person S just when S is more justified in believing the proposition than he or she is in believing its negation.[36]

If the conception of being justified that Chisholm employs is supposed to be closely related to truth, I would reject CP right away. There is surely no good reason to suppose that if anyone accepts something that is not disconfirmed by his or her total evidence, that proposition is apt to be true or even close to the truth. Ignorant and barbarous people, as Hume would say, believe all sorts of patently false things that are not disconfirmed by the evidence available to them.[37] Surely things so believed are not true more often than not. People pertinaciously obstinate in their delusions (another Humean turn of phrase) refuse to consider evidence ostensibly contrary to those delusions, so they are, in effect, insulated from anything that might disconfirm them. The mere fact that they are believed is hardly evidence in their favor.

Reasonable people who have a sincere interest in discovering the truth will not protect their illusions this way, so the fact that their beliefs are not disconfirmed by the evidence available to them is an epistemically much more significant fact. May we not suppose that such people have probably confronted ostensibly disconfirming evidence for their beliefs and ruled it out on rational grounds? May we not conclude that their surviving beliefs are apt to be true more often than not? I would say no; a general conclusion of this kind is excessively indiscriminate. A well-considered judgment on this matter must take into account contingent

---

[34] Ramsey (1931), p. 182.
[35] Chisholm (1989), p. 72.
[36] *Ibid*., p. 87.
[37] Hume thought of some human beings as "extremely ignorant and stupid" and "ready to swallow even the grossest delusion." See Hume (1777), p. 120.

facts about these people and about the kind of beliefs they are apt to form. If, like the investigators Locke would have commended, they are in the habit of proportioning their beliefs to the evidence, they are unlikely to have beliefs for which they lack positive supporting evidence; their non-disconfirmed beliefs are probable only because they are rendered so by such evidence. Other people might be slightly less circumspect in forming beliefs, and some might not be circumspect at all. The latter may be willing to consider contrary evidence but they may readily form beliefs about domains for which there is little or no possibility of obtaining evidence: they may have elaborate theologies, mythologies, or fanciful histories that dominate their thoughts but have no testable consequences. The otherworldly beliefs that survive disconfirmation for these people may be almost invariably false; there is no reason to suppose that they are ever true.

Although Chisholm's unqualified principle CP is thoroughly unacceptable,[38] a qualified version strikes me as defensible. The first qualification concerns the word "acceptance." This is really not a good word for a properly qualified version of CP, because the notion of probability appropriate for such a principle is the degree of certainty notion I have been discussing, and the degree of probability appropriate to the "accepted" hypothesis must be minimal if the subject lacks positive evidence. For a properly qualified CP, the acceptance in question is best described as that of *weak acceptance on a trial basis*. The second qualification is closely related to the first one: the word "probable" must refer to a minimal degree of certainty. Since 0.5 represents probabilistic indifference, where the certainty of "P" is the same as the certainty of "~p," and 0.75 represents moderate certainty, the midpoint between indifference and certainty, minimal probability can be taken to be somewhere between 0.55 and 0.6. A probability in this range can be corrected by further testing involving Bayes' theorem. Its posterior value will depend on the way it is supported empirically.

I have now explained how the probabilities needed to apply Bayes' theorem can justifiably be obtained. The priors needed for hypotheses are obtained by testing more primitive priors that were initially adopted as conjectures having predictive (Lycan would say "explanatory") potential; likelihoods are obtained principally by deduction from hypotheses and auxiliary assumptions; and priors needed for evidence statements are deduced from background information concerning the theoretical principles applicable in the case at hand. The results of predictions are ascertained, ultimately, by observation. In practice observations involve the application of theoretical knowledge, but this knowledge ultimately results from the observation of predicted results. When we update priors by applying the rule of conditioning, we are in effect assigning a probability of 1 to a verified observation statement. A lesser probability is, strictly speaking, appropriate, since empirical statements are always uncertain to some degree.[39] In practice this uncertainty is commonly disregarded, since precise probabilities are not generally required for most empirical investigation. We can always be more precise if we think we have to.

There are many issues pertinent to Bayesian reasoning that I have not touched on here; in fact, there are problems of varying seriousness that specialists in probability theory continue to debate.[40] One problem concerns old evidence. I observed that hypotheses are not strongly supported by the occurrence of events with a high prior probability, but sometimes there is a serious question about how old or familiar data is to be explained, and a hypothesis' ability to account for that data is assumed to count strongly in favor of it. Is such an assumption always objectionable? I would say no. If there is a serious question about how some familiar data is to be explained—if no available explanatory account is accepted as applying to it—then that data can reasonably be considered unlikely in relation to accepted principles, and statements reporting its occurrence can be assessed as such for explanatory purposes. Theoretically, the data, though actual and therefore probable, is surprising; and its surprise value is what is represented by the low probability value. A plausible hypothesis that can predict its occurrence is therefore increased in probability value; for purposes of the calculation its actuality is ignored. When the calculation is completed, the fact that the data is obtained may then be used to update the probability value of the hypothesis, making its posterior value a new prior. It seems to me that this procedure raises no significant problems. An investigator who supports a hypothesis by old data will of course want to test it further by means of new data, but the old data is more significant than it would be if it had to be assigned a probability close to one for purposes of the calculation.

Although there is much more that needs to be said about Bayesian inference,[41] what I have said is enough to show that it is far preferable to Inference to the Best Explanation, the alternative

[38] For a criticism of Chisholm's defense of CP see Appendix 5.
[39] See the helpful discussion in Rosenkrantz (1981), 3.6, 3-4.
[40] For discussion see Earman (1992) and the very perceptive review of Earman in Hájek and Skyrms (2000).
[41] See Talbot (2001), Howson and Urbach (1989), Skyrms (1986), Rosenkrantz (1981), and Phillips (1973).

widely acceptable today. Apart from the actual defects that I have noted, IBE has no evident rationale in the first place: even lacking identifiable defects of the kind I mentioned, there is no evident reason why a rational person should be moved to adopt it. The same is not true of Bayesian inference. Bayes' theorem is a mathematically sound principle, inferable from axioms that, interpreted in the way I have suggested, are reasonably regarded as analytic.[42] This gives Bayesian inference a rationale not possessed by other principles of experimental inference. Of course, the results of a Bayesian inference are not analytic, for nonanalytic premises are required for the inference. But these nonanalytic premises are rationally defensible for the reasons I have given. Nothing analogous appears to be true for Inference to the Best Explanation, the only evident alternative for reasoning about matters that transcend the domain of the observable.

### The BIV Hypothesis Again

If Bayes' theorem provides a valid form for experimental inference and if, in addition, we may justifiably accept (under the conditions I mentioned) observational premises, prior probability assignments, and delimited sets of alternative hypotheses, the way is then open for confirming the facts about observers that support or refute what they claim to have observed or remembered. More than this is actually possible: Means are now available for dismissing Putnam's BIV hypothesis, and doing so by straightforward empirical means. Since this latter task is more fundamental than that of supporting pertinent facts about observers or those purporting to remember something, discussing the strategy for dismissing the BIV hypothesis is a fitting way of ending this chapter.

To show that Putnam's hypothesis deserves to be rejected on empirical grounds, we need only show that an alternative, realist hypothesis—one formulating our best estimate about the actual nature of ourselves and our world—is better supported, empirically, than the hypothesis about BIVs and their relation to the computer that orchestrates their delusional experiences. This would seem to be a very easy thing to so, because the BIV hypothesis represents a mere conceptual possibility, one that, like Descartes' hypothesis of an evil genius, has absolutely no evidence in its favor. If we can show that the realist hypothesis we actually accept has *any* evidence in its favor, we would evidently accomplish this task in short order. Unfortunately, the issue is not quite this simple. If we take the BIV hypothesis seriously as even a possibility, we will be thinking of our available evidence as consisting of facts about the subjective experience of the relevant subjects. So the question for us to consider is whether, by reference to that kind of experience, we can show that our realist view is better supported than Putnam's BIV hypothesis.

To support the view that we actually accept—it sounds odd to call it a hypothesis, but I shall do so for the sake of argument—it is convenient to begin with a simpler, more specific hypothesis, ST, one that entails some of the consequences of the accepted alternative for my own current physical situation in the world. ST asserts something very specific--namely, that I am (really) sitting at my desk in my study looking into my word processor and that there is behind me a Bertoia chair covered with red upholstery. As I implied at the beginning of this section, Bayes theorem permits us to show that this simple hypothesis can readily be confirmed by reference to my subjective experience. The procedure is straightforward.

Given ST and some related assumptions concerning the nature of what I take sitting, looking, seeing chairs, and acting to be, I can justifiably predict that if I will to turn around 180° and subsequently have the experience of doing so, I will in fact turn around this way and subsequently see, and so have the experience of seeing, the red Bertoia chair. I so will and I have the predicted experiences. The likelihood here, the probability of having the indicated experiences on the supposition that the hypothesis and auxiliary assumptions are true, can reasonably be set very high: the hypothesis and assumptions can be elaborated in a way that warrants this prediction with a high degree of certainty. The antecedent probability[43] of having the experience of seeing a red Bertoia chair if I merely will to turn around can, by contrast, be reasonably considered quite low, since willing to turn my head is rarely followed by such a visual experience: it is so followed only in cases when I believe I am in my study looking into my word-processor. What then about the prior probability of my hypothesis, ST? If I am not to beg the question against BIV, I cannot make use of the evidence I would normally advance if I if were asked to defend my conviction that I am at my desk in my study facing my word processor. That evidence is based on defeasible presumptions that are now in question. Without that evidence, I cannot support a high prior probability for ST; I can only give it a moderately low prior probability in accordance with the strategy I described when I mentioned Chisholm's principle

[42] See Appendix 6.
[43] This probability is actually conditional on the assumptions about sitting, looking, seeing chairs, and acting that I have mentioned. I am here using the more complicated version of Bayes' theorem set forth in footnote 27.

CP. But even with a moderately low prior for ST, a high value for P(e/ST) and a low value for P(e) yields a high value for P(ST/e) and, in view of the truth of e, a high posterior value for P(ST).

Of course, since the prior for ST was assumed on a trial basis, further tests are required, possibly with slightly different priors chosen for ST. But if the predications made on the basis of these different priors are satisfied to a comparable degree (as they would be, since the predictions would be the same), the posterior values for ST would remain high and justify a high value for that hypothesis. Since BIV is incompatible with ST, a high posterior value for ST requires a low value for BIV. Thus, BIV is effectively ruled out in favor of ST.

A supporter of BIV might object to this reasoning, saying that the antecedent probability of BIV deserves to be set as high (for purposes of the argument) as ST and that it would be confirmed just as strongly as ST since it will warrant the same predictions as ST. This objection fails, however. It is true that on a generous interpretation of what BIV asserts—one specifying appropriate links between the intentions and know-how of the scientific maniacs, the computer program, and the experiences of the disembodied brains—BIV will predict everything that ST predicts. But BIV makes many untestable claims that ST (or the realist hypothesis that ST represents) does not make, and this surplus content requires it to have a lower prior probability than ST—far lower, in view of its extravagantly rich sci-fi content.

Why does the surplus, untestable content of BIV require it to have a lower prior probability than ST? The reason is this. If BIV has every testable consequence that ST has but not vice versa, then BIV can be divided into two parts, A and B, one of which, A, represents the untestable content of BIV and the other, B, represents the part that is empirically equivalent to ST. Now A is probabilistically independent of B; P(A on B) = P(A). (If A and B are independent, the probability of either is unaffected by the truth of the other.) We knows that A and B are independent because B represents the part of BIV that does not have testable consequences, and A represents the rest of it, which has the testable consequences of the fully testable RT. But if A and B are independent in this way, $P(A \wedge B)$ is less then both P(A) and P(B) if A and B are neither analytically true nor analytically false (as they are in the BIV case). This is a consequence of the theorem for the probability of such conjunctions: If A and B are independent, $P(A \wedge B) = P(A) \times P(B)$.[44] Thus, owing to its untestable part B, the prior probability of BIV is less than the prior probability of ST. Given Bayes' theorem, we can therefore conclude that ST is more strongly confirmed by its supporting evidence than BIV is by its supporting evidence. By hypothesis, the supporting evidence in both cases is the same.

It is vital to realize that simplicity is not the only consideration that matters here. Equally important is the fact of empirical equivalence, the fact that the rival theories have the same testable consequences. Someone might argue that the ST hypothesis may be simpler, in the indicated way, than the BIV hypothesis, but that other rivals to ST may be even simpler. One such rival is a form of phenomenalism, the doctrine that only experiences are fundamentally real and that words such as "person" and "physical body" refer to nothing other than aggregates of, or constructs wholly reducible to, experiences.[45] There is no doubt that a theory of this kind is ontologically simpler—simpler in the sense of postulating fewer irreducible objects—but no such theory has ever been worked out in a satisfactory way,[46] and the best examples of the kind have been acknowledged to be inadequate by their authors.[47] Apart from this, the testable consequences of such a theory are by no means the same as those of ST: the latter, even the limited representative of the full-blown realist hypothesis that we actually accept, concerns the experiences of this or that person, and a person is a subject of experiences, not an aggregate of them.

I do not wish to plunge into the swamp of suggestions and replies that were once pursued when "Our Knowledge of the External World" was the leading topic on a philosopher's agenda. My argument against the BIV hypothesis was prompted by an imagined challenge--that of showing that skeptical hypotheses that are possible alternatives to the realist views we normally accept are not uncritically brushed aside but deserve to be rejected for identifiable reasons. I have tried to put my finger on some of those reasons. Much more could be said; the ST hypothesis as I described it was not the full-blown realist alternative to the BIV one: it was merely a limited hypothesis about me in my study, not a full-scale hypothesis about the nature of the world I inhabit and the creatures that share it with me. To move from ST to the alternative I have

---

[44] When two fractions are multiplied together, the product is less than either fraction. $1/2 \times 1/2 = 1/4$.

[45] Fumerton (2005) suggests that Berkeley's hypothesis about God would be a simpler hypothesis, but phenomenalism would be even simpler and would not involve the enormous untestable surplus of Berkeley's hypothesis.

[46] I outline my objections to phenomenalism in Aune (1991), chapter four, section 8.

[47] Bertrand Russell gave up on phenomenalism, which he had developed in *Our Knowledge of the External World* (1914), as early as 1927, when he published his *Analysis of Matter*; Carnap abandoned the phenomenalist program of his *Logische Aufbau der Welt* (written 1922-25 but published in 1928) as early as 1936, when he published "Testability and Meaning." For discussion, see Freidman (1999), ch. 5.

inadequately sketched, I should defend more of the presumptions we commonly make in thinking about the world. One is that the animated bodies I call people do not just behave intelligently and smile or frown when they are stroked or poked, but also think and feel much as I do. This presumption, also high on the philosophical agenda at another time, can be defended as well by a Bayesean strategy, but I have no interest in pursuing it here.

Historically, philosophers presented with skeptical hypotheses of the kind I have been considering have tried to refute them—to rule them out—by some kind of a priori strategy. They are cognitively meaningless; they fall short of the requirements for objective reference; they presuppose a nonsensical private language; they violate the sound requirements for the acceptable interpretation of meaningful discourse; and so forth. But according to the standards of a reasonable empiricism, these hypotheses are clearly meaningful. We know exactly what they mean; they would not be problematic for us if we could not understand them. They are, of course, far fetched, but that is not enough to show that they are false. What entitles us to reject them is that they are not nearly as well supported by available evidence as the hypotheses we accept. This evidence supporting accepted hypotheses is not perfect; it does not render them acceptable beyond the shadow of a theoretical doubt. Yet it is sufficient for knowledge in the sense that we commonly employ. If we recognize only perfect knowledge, we will have to cope with a form of skepticism.[48] But there is no actual need to proceed this way.

When I hear philosophers seriously endorsing a skeptical view of human knowledge, I think of a photograph I recently saw of the Martian landscape, one of a series taken by the Rover vehicles in 2004 and transmitted back to earth. The photo looked a lot like a photo of the Mojave Desert. The color of the landscape was different, but it seemed very similar nevertheless. When the photo comes to mind, I am struck by the extraordinary achievement it represents. The engineers who created the Rover vehicles and the rockets that carried them to Mars had to possess an enormous amount of detailed knowledge about a bewildering variety of phenomena, and those who carried out the missions that produced the photos had to be right in more calculations than I could possibly enumerate. Yes, there are many things we do not know about our universe and ourselves, but there is an astonishing amount that we do know very well. It is sometimes hard to believe that the creatures who were hunting with arrows and spears ten centuries ago can now send robots to distant planets and then later leisurely view in their home television screens the photos sent back to them, doing so as they sip a cup of coffee or drink a glass of wine.

## Concluding Remarks

The empiricist epistemology I have defended in this book is partly classical, partly reformed. Its basic structure is largely classical, recalling the empiricism of David Hume. Corresponding to Hume's division of the objects of human inquiry into "relations of ideas" and "matters of fact and existence," I have defended a distinction between analytically true and synthetically true statements. My distinction is not categorical, however; it does not place every true statement into one of two disjoint classes. Natural dialects or even idiolects are not sufficiently determinate to allow such a distinction, but careful speakers seriously concerned about the precision and truth-value of what they say can make their meaning sufficiently determinate to place the important things they want to say into one or the other of these classes.[49] Assertions that are analytically true are either logically true or A-true in Carnap's sense. This conception of analyticity rests on a distinction between the theorems of some assumed system of formal logic and assertions warranted by stipulative explications or determinate verbal conventions. The rationalist's supposed example of a synthetic a priori truth, ""Nothing can be both determinately green and determinately yellow all over," is actually an assertion of this last kind; it is in fact inferable, as I show, from a verbal convention about what counts as a single determinate color. Spelling out all the distinctions and qualifications needed to extend an analytic/synthetic distinction to thoughts as well as statements requires a fairly elaborate exposition; what I have just said is merely a skeleton summary of points I defended in chapters two and three.

Another classical element of my empiricism is the list of basic sources by which we can ascertain matters of fact and existence. Hume referred to these sources as observation, memory, and experimental inference. Each source is beset by distinctive problems, but each is unquestionably a source of genuine knowledge, at least when measured by ordinary, imperfect standards. A key component in all three sources is human experience, which is fundamentally directed to a world of things and persons. It is by hearing my fiddle that I know it to be in or out of tune; it is by tasting my wine that I know it to be Zinfandel, and it is by looking at a clock that I know what the time is supposed to be. The knowledge of the world that we get from the three

[48] As Fogelin (1994) in effect argues.

[49] Mathematical assertions are possible exceptions; see my chapter two, p. 74f and footnote 2.

sources is not certain, but it can be improved by other data. Initial probabilities are transformed into progressively more certain convictions. Because empirical confirmation is a dynamic process, taking place in many different directions, the totality of what we know at any particular time is inevitably somewhat disorganized; we can only strive for greater unity. As a consequence of these facts, popular depictions of the structure of what we know—foundationalism, coherentism, or even Quine's holism—are all inaccurate in some significant way.

As I have emphasized, empirical knowledge as we commonly understand it is not the kind of knowledge that Descartes was after. It falls far short of rational certainty. When knowledge is understood this last way, far-fetched possibilities such as Descartes' evil demon or Putnam's brains in a vat assume an epistemic importance they do not really deserve. We have no way of *proving* that these possibilities are not actual—that the stories they involve are in fact false—but we can show that they are far-fetched and that the contrary views we actually accept are much more likely to be true. And this is enough for a reasonable empiricism, one appropriate for a philosopher who aspires to be tough-minded but epistemically up-to-date. Classical empiricists often emulated Descartes in their quest for certainty, and skeptical scenarios have therefore persistently threatened them. We can avoid this outcome by a more realistic estimate of the kind of certainty we can hope to achieve.

Two assumptions once thought distinctive of a responsible empiricism must be firmly set aside. One is the assumption that our empirical knowledge or well-founded opinion must rest on a foundation of subjective experience. Not only does our empirical knowledge fail to rest on anything that deserves to be called a foundation, but the nature of our subjective experience is also, as I noted, quite questionable, generating on-going controversy among philosophers and even empirical scientists. The other objectionable assumption is that inherently unobservable objects are unknowable and cannot meaningfully be described or referred to. The classical view of meaning and legitimate reference on which this assumption is based is simply untenable. Meaningful words or ideas need not, as Hume said, be "traced back to original impressions";[50] they arise from inbuilt neural mechanisms and goal-directed, usually cooperative behavior,[51] neither of which is understandable by reference to immediate impressions.

Because the use of Bayes' theorem provides a rationally acceptable means of confirming hypotheses about objects and process that fall outside the domain of the observable, knowledge as we commonly understand it need not be restricted to observable phenomena. Saying this does not commit me to the view that acceptable scientific theories must always be interpreted "realistically," as describing actually existing, mind-independent objects. Acceptable theories can do different things: some can provide mere models or vehicles of prediction, which do not purport to describe anything observable. The most that we can generally require of an empirical theory is that it be, in van Fraassen's words, "observationally adequate."[52] But a general requirement of this kind does not preclude a realistic interpretation of some theories. That would be going too far. Empirical reality can contain many things too remote or too small to be perceived by us, and it can contain numerous things that we could never observe for reasons that we cannot even anticipate. A reasonable empiricism prescribes experience as our ultimate basis for empirical knowledge, but it does not limit our knowledge to things that we can observe or otherwise experience.

---

[50] Hume, *Enquiry*, section II.
[51] The support for this is now given by scientists rather than philosophers; for a general discussion see e.g. Pinker (1994).
[52] van Fraassen (1980).