CHAPTER FOUR

Duhem, Kuhn, and Bayes

DEFENDERS of the Bayesian Way can and do argue that even if scientists are not conscious or unconscious Bayesians, reconstructing scientific inference in Bayesian terms is of value in solving key problems of philosophy of science. In this chapter I will consider how Bayesian reconstructions have been used to grapple with Duhem's problem, and to bridge the logical empiricist approach to confirmation with the historicist approach promoted by Kuhn. In both cases I will argue that if the goal is solving rather than reconstructing problems, then the Bayesian Way comes up short.

4.1 THE BAYESIAN WAY OUT OF THE DUHEM PROBLEM

The problem for which the Bayesian Way is most often touted as scoring an impressive success is the Duhem problem—the problem of which of a group of hypotheses used to derive a prediction should be rejected when experiment disagrees with that prediction. Although I will argue that the Bayesian Way out of Duhem's problem is really no way out at all, my aim is not primarily negative. Rather, my hope is to lay the groundwork for a satisfactory non-Bayesian approach to the problem based on error statistics.

Some philosophers of science dismiss the Duhem problem as the product of old-fashioned (hypothetico-deductive) philosophy of science and therefore not really an issue for New Experimentalists. What Duhem's problem shows, strictly speaking, is that logic alone permits an anomalous result to be blamed not on the primary hypothesis being tested, but on the host of auxiliary principles and hypotheses involved in testing. And we know formal logic is not all we have at our disposal. But the problem that still remains is to show that there are good grounds for localizing the bearing of evidence. If an inference account cannot at least make headway toward showing which assignment of error is warranted, it cannot be seen to have gotten around the Duhem problem in its modern guise.

Lakatos, we saw, attempted to improve on Popper in the light of

Duhem's problem as brought home by Kuhn. For Lakatos, anomalies are blamed on suitable auxiliary hypotheses, hard-core theories remaining protected. But he conceded that any hard-core theory can be defended "progressively" this way. Bayesians believe that they have a more adequate solution to Duhem's problem, that "the questions left unanswered by Lakatos are answered with the help of Bayes's theorem" (Howson and Urbach 1989, 96). They and other Bayesians appeal to the Bayesian strategy of Jon Dorling (1979), which I will outline shortly. In the section "The Duhem Problem Solved by Bayesian Means" (p. 96), Howson and Urbach declare just that. Let us see what they mean.

The Duhem Problem Solved by Bayesian Means

When Bayesians say they can solve Duhem's problem, what they mean is this: Give me a case in which an anomaly is taken to refute a hypothesis H out of a group of hypotheses used to derive the prediction, and I'll show you how certain prior probability assignments can justify doing so. The "justification" is that H gets a low (or lower) posterior probability than the other hypotheses. As with the general Bayesian Way of explaining a scientific episode, solving Duhem comes down to a homework assignment—not to say a necessarily easy one—of how various assumptions and priors allow the scientific inference reached to be in accord with that reached via Bayes's theorem.

In addition to accounting for specific episodes, the Bayesian Way can be used to derive a set of general statements of the probabilistic relationships that would have to hold for one or another parceling out of the blame. These equations are neat, and the algorithms they offer for solving such homework problems are interesting. What they do not provide, however, is a solution to Duhem's problem. Duhem's problem, as Howson and Urbach themselves say, is to determine "which of the several distinct theories involved in deriving a false prediction should be regarded as the false element" (Howson and Urbach 1989, 94). The possibility of a degree of belief reconstruction does not help to pinpoint which element ought to be regarded as false.

From all we have already seen, we might expect the subjective Bayesian to retort that I am misunderstanding the subjectivist account. For the subjective Bayesian, the hypotheses an agent ought to consider disconfirmed *are* the ones with low posterior probabilities, and these follow deductively from the agent's prior degrees of belief (and other subjective probabilities), which agents are assumed to have. That is what a subjectivist means by an inference being rational. Dorling (1979), to his credit, admits as much. He says that adopting a personalist reconstruction "automatically" yields a resolution of Duhem, but

quite correctly stresses "that it is the adoption of a *personalist* Bayesianism which yields this way out of the Duhem problem" (p. 178). The question that remains is whether to adopt the Bayesian Way out is really to have a way out of the problem. Not, I claim, if the problem is understood normatively. What the Bayesian Way offers, at best, is a way of reconstructing given inferences probabilistically. The Duhem problem, if it is not simply defined away, just returns as the problem of justifying the correctness of the probabilities in the Bayesian equations.

Since Dorling's work is credited as the exemplar for the Bayesian solution to Duhem, I will take it as my example too.

Dorling's Homework Problem

Dorling considers a situation where despite the fact that an anomalous result e' occurs, the blame is placed on an auxiliary hypothesis A while the credibility placed on theory T is barely diminished. In Dorling's simplified problem, only one auxiliary hypothesis A is considered (I am replacing his H with A).

In the historical case considered here, Dorling (1979, 178) takes T to be "the relevant part of solidly established Newtonian theory which Adams and Laplace used" to compute e, the predicted secular acceleration of the moon, which conflicted with the observed result e'. The auxiliary, A, is the hypothesis that the effects of tidal friction are not of a sufficient order of magnitude to affect appreciably the lunar acceleration.

Dorling's homework problem is to provide probability assignments so that, in accordance with the episode, an agent's credibility in theory T is little diminished by the anomaly e', while the credibility in auxiliary A is greatly diminished. We can sidestep the numerical gymnastics to get a feel for one type of context where the agent faults A. Afterwards I will give a numerical algorithm (calculated at the end of section 4.1).

Theory T and auxiliary A entail e, but e' is observed. When might e' blame A far more than T? Here's one scenario sketched in terms that I intend to be neutral between accounts of inference. Suppose (1) there is a lot of evidence for theory T, whereas (2) there is hardly more evidence for the truth of auxiliary hypothesis A than for its falsity. Suppose, further, that (3) unless A is false, there is no other way to explain e'. This is a rough account, it seems to me, of a situation where e' indicates (or is best explained by) A being in error.[1]

1. A more extreme situation would give a very low prior probability to hypothesis A. Dorling is trying to describe a case where it is not so obvious how things come out.

A Bayesian rendering may be effected by inserting "agent x believes that" prior to assertions 1, 2, and 3. We then have a description of a circumstance where the agent believes or decides that A is discredited by e'. Nothing is said about whether the assignments are warranted, or, more importantly, how a scientist should go about determining where the error really lies. Assigning the probabilities differently puts blame elsewhere, and the Bayesian "solution" is not a solution for adjudicating such assignments.

The Numerical Solution to the Homework Assignment

The numerical "solution" that corresponds to what I described above is this: The scientist's degree of belief is such that a high degree of belief is initially accorded to T (e.g., $P(T) = .9$); in any case, it is substantially more probable than A, which is considered only slightly more probable than not (e.g., $P(A) = .6$). These numbers are introduced by the personalist, Dorling explains, as approximate descriptions of the belief state of a particular scientist at the time. Let us see how we might describe the agent's beliefs so that the third and key assumption is cashed out probabilistically.

First, imagine the agent considering the possibility that auxiliary hypothesis A is true:

The agent contemplates auxiliary A *being true.* Clearly, T could not also be true (since together they counterpredict e'). But might not some rival to T explain e'? Here is where the key assumption enters. The agent believes there to be no plausible rival that predicts e'. That is to say, the agent sees no rival that, in his or her opinion, has any plausibility, that would make anomaly e' expected. In subjective probability terms, this becomes

a. The probability of e', given that A holds and T is false, is very small. Let this very small value be ε.

Since the anomaly e' has been observed, it might seem that the agent would assign it a probability of 1. Doing so would have serious ramifications (i.e., this is the "old-evidence problem"). To avoid assigning degree of belief 1 to e', Bayesian agents need to imagine how strongly they *would have believed* in the occurrence of anomaly e' *before* it was observed—no mean feat. But never mind the difficulties in assigning such probabilities just now (see chapter 10). The Bayesian assumes that the agent can and does make the key assumption that, in the agent's view, the e' observed is extremely improbable if A is true. Now consider the agent's beliefs assuming that auxiliary A is false.

The agent contemplates auxiliary A *being false.* In contrast, if auxiliary A were false, the agent would find e' much more likely than if T were

false and A true. In fact, Dorling imagines that scientists assign a probability to e', given that A is false, 50 times as high as that in (a), whether or not T is true. That is, $P(e' \mid A \text{ is false}) = 50\varepsilon$. We have

b. i. The probability of e', given that T holds and A is false, is 50ε.
ii. The probability of e', given that T is false and A is false, is 50ε.

Of course, (i) and (ii) need not be exactly equal, but what they must yield together is a probability of e' given A is false many times that in (a). A further assumption, it should be noted, is that T and A are independent.

Together, (a) and (b) describe a situation where the outcome e' is believed to be far more likely if A is false than if A is true. This yields assumption 3. The result is that the posterior probability of T remains rather high, that is, .897, while the posterior of A becomes very low, dropping from .6 to .003.

This gives one algorithm—Dorling's—for how evidence can yield a Bayesian disconfirmation of auxiliary A, despite A's being deemed reasonably plausible at the start. Nonquantitatively put, the algorithm for solving the homework problem is this: Start with a suitably high degree of belief in T as compared with A, believe no plausible rival to T exists that would make you expect the anomalous result, and hold that the falsity of A renders e' many times more expected than does any plausible rival to T.

Reconstructing versus Solving Duhem

Dorling's homework problem can be done in reverse. Scientists who assign the above degrees of belief, but with A substituted for T, reach the opposite conclusion about T and A. So being able to give a Bayesian retelling does not, by itself, say which apportionment of blame is warranted.

Bayesians may retort that the probabilities stipulated in their reconstruction are plausible descriptions of the beliefs actually held at the time, and others are not. That may well be, though it is largely due to the special way in which they describe the prediction. I leave that to one side. For my own part, I have no idea about the odds "a typical non-Newtonian would have been willing to place [on] a bet on the correct quantitative value of the effect, in advance even of its qualitative discovery" (Dorling 1979, 182). (Something like this is the contortion required to get around assigning e' a probability of 1.)

Nor is it easy to justify the prior probability assignments needed to solve the homework problem, in particular, that theory T is given a prior probability of .9. The "tempered personalism" of Abner Shimony

(e.g., 1970) advises that fairly low prior probabilities be assigned to hypotheses being considered, to leave a fairly high probability for their denial—for the "catchall" of other hypotheses not yet considered. The Dorling assignment leaves only .1 for the catchall hypothesis.

A Highly Qualified Success? If Bayesian reconstructions fail to count as solving Duhem, it seems fair to ask what value such reconstructions might have. Bayesians apparently find them useful. John Earman, for example, shares my position that the Bayesian Way is no solution to Duhem. While calling it a "highly qualified success for Bayesianism," Earman finds that "the apparatus provides for an illuminating representation of the Quine and Duhem problem" (Earman 1992, 85). For my part, I find the problem stated by Duhem (1954) clear enough—how to determine the error responsible:

> The only thing the experiment teaches us is that among the propositions used to predict the phenomenon . . . there is at least one error; but where this error lies is just what it does not tell us. The physicist may declare that this error is contained in exactly the proposition he wishes to refute, but is he sure it is not in another proposition? If he is, he accepts implicitly the accuracy of all the other propositions he has used, and the validity of his conclusion is as great as the validity of his confidence. (Duhem 1954, 185)

This last clause can be put in Bayesian terms by replacing "the validity of his confidence" with "the validity of his prior and other degree of belief assignments." But I do not see how attaching a degree of belief phrase to the claims in Duhem's statement helps to illuminate the matter. Indeed, attaching probabilities to statements only complicates things.

But there is something that might be said about the Bayesian reconstructions that may explain why philosophers find them appealing to begin with. A purely syntactical theory of confirmation along the lines of a hypothetico-deductive account seems to lack a way to account for differential assignments of blame for an anomaly.[2] Two different cases may go over as the same syntactical configuration, even though our intuitions tell us that in one case the primary hypothesis is discredited while in the other the auxiliary is. The complaint against syntactical approaches is correct. But this shows only that syntax alone won't do and that substantive background knowledge is needed. What

2. Even Glymour's bootstrapping version, Earman argues, seems to have no way to solve it.

it does not show, and what I have been urging we should deny, is that the background should come in by way of subjective degrees of belief.

A Sign of Being a Correct Account? Howson and Urbach follow Dorling's treatment in their own example, giving assignments very similar to, though less striking than, Dorling's. The ability of the Bayesian model to accord with actual cases of attributing blame, they conclude, shows that "Bayes's Theorem provides a model to account for the kind of scientific reasoning that gave rise to the Duhem problem" (Howson and Urbach 1989, 101). If this just means that there are Bayesian reconstructions of the sort we have been considering, then we can agree. However, Howson and Urbach go on to claim that the ability to give a Bayesian reconstruction of cases shows "that the Bayesian model is essentially correct"! (p. 101). But merely being able to offer reconstructions of episodes says nothing about the Bayesian model's correctness.[3]

If the name of the game is reconstruction, it is quite simple to offer a non-Bayesian one. How would our error-testing model reconstruct an episode where an auxiliary A is blamed, rather than theory T? We would want to distinguish between two cases: (*a*) the case where there are positive grounds for attributing the error to auxiliary A, and (*b*) the case where there are simply inadequate grounds for saying an error in A is absent. In coming out with a posterior probability of .003 in A, Dorling is describing the case as (*a*), yet anomaly e' itself seems at best to warrant regarding it as in case (*b*)—where there is simply not enough information to attribute blame to T.

An error-statistical description of the episode might go like this: Theory T is not shown to be in error as a result of anomaly e' unless the evidence warrants ruling out the possibility that an error in auxiliary hypothesis A is responsible. Evidence does not warrant ruling out an error in auxiliary A unless A has been shown to pass a sufficiently severe test. But the assumption of lukewarm evidence for A (reconstructed by Dorling as its having a prior probability of .6) would be taken as denying that A had passed a severe test. This explains why e' was not taken to discredit T. To take e' as grounds for condemning T would be to follow a very unreliable procedure. To reconstruct an episode as a case of (*a*), in contrast, would require there to be positive grounds to consider A false, and its falsity to blame for the anomaly. In that case what must have passed a severe test is hypothesis "not-A":

3. Similar criticisms of the Bayesian solution to Duhem are raised by Worrall (1993).

that the extraneous factor (tidal friction) is responsible for the anomalous effect (lunar acceleration).

My error statistics reconstruction enjoys several advantages over the Bayesian one: First, it does not suppose that for any anomaly there is some inference to be reached about where to lay the blame. The description in (*b*) may be all that would be allowed until positive grounds for fingering an auxiliary were obtained. Second, the question whether there are grounds for an error in *A* does not turn on opinions in *T* and there is no need to imagine having a prior in all the other possible theories (i.e., the so-called catchall). This second reason leads to a third, which is what allows us to go beyond mere reconstruction: unlike the probabilities needed for the Bayesian reconstruction, philosophers do not have to invent the components we need in depicting the scientific inference, nor work with make-believe calculations (e.g., imagining the odds scientists would place if they did not already know the evidence).

There seems to be no suggestion, even by Bayesians, that scientists actually apply Bayes's theorem in reaching their conclusion. Most important, the Bayesian description fails to capture how Duhemian problems are actually grappled with before they are solved. Adjudicating disputes with a measure of objectivity calls for methods that can actually help to determine whether given auxiliaries are responsible for the anomaly. Scientists do not succeed in justifying a claim that an anomaly is due to an auxiliary hypothesis by showing how their degrees of subjective belief brought them there. Were they to attempt to do so, they undoubtedly would be told to go out and muster evidence for their claim, and in so doing, it is to non-Bayesian methods that they would turn.

What's Belief Got to Do with It?

Howson and Urbach (1989) state, without argument, that "by contrast [with the Bayesian model], non-probabilistic theories seem to lack entirely the resources that could deal with Duhem's problem" (p. 101) where "non-probabilistic theories" include the error statistics methods of Fisher and Neyman and Pearson. In truth, these methods contain just the resources that are needed and regularly relied upon to solve real-life Duhemian problems.

A major virtue of the error statistics approach is that the issue of whether the primary or auxiliary hypothesis is discredited is not based on the relative credence accorded to each. The experiment is supposed to find out about these hypotheses; it would only bias things to make interpreting the evidence depend on antecedent opinions. After all, in

Dorling's examples, and I agree that the assumption is plausible, theory *T* is assumed to be *independent* of auxiliary *A*. There is no reason to suppose that assessing auxiliary *A* should depend at all on one's opinion about *T*. What is called for are separate researches to detect whether specific auxiliaries are responsible for observed anomalies.

Let me allude to an example to be considered later (chapter 8). When one of the results of the 1919 eclipse experiments on the deflection of light disagreed with Einstein's prediction, there was a lengthy debate about whether the anomaly should be attributed to certain distortions of the mirror, to Einstein's theory, or to something else. The debate over where to lay the blame was engaged in by scientists with very different opinions about Einstein's theory. Such attitudes were no part of the arguments deemed relevant for the question at hand. The relevant argument, put forth by Sir Arthur Eddington (and others), turned on a rather esoteric piece of data analysis showing (holdouts notwithstanding) that the mirror distortion was implicated.

Eddington believed in the correctness of Einstein's account, but nobody cared how strongly Eddington believed in Einstein. Quite the contrary—it only made those who favored a Newtonian explanation that much more suspicious of Eddington's suggestion that the faulty mirror, not Einstein's account, was to blame. Being an ardent proponent of either of the two rivals entered the debate: it explained the lengths to which players in the debate were willing to go to scrutinize each other's arguments. But ardor did not enter into the *evidential appraisal* of the hypotheses involved.

The argument to blame an auxiliary such as the mirror is the flip side of the argument to rule out an artifact. Here the anomalous effect may be shown to go away when there is no distortion of the lens. Additional positive arguments that the lens was the culprit were given, but I will save those for later.

Ronald Giere (1988) suggests a "technological fix for the Duhem-Quine problem" (p. 138), observing that often auxiliary hypotheses are embodied in instruments, and "Scientists' knowledge of the technology used in experimentation is far more reliable than their knowledge of the subject matter of their experiments" (p. 139). My position for solving Duhem extends this technological fix to include any experimental tool. It is the reliability of experimental knowledge in general, the repertoire of errors and strategies for getting around them, that allows checking auxiliaries, and for doing so quite apart from the primary subject matter of experiments.

When it comes to finding out which auxiliaries ought to be blamed, and to adjudicating disputes about such matters, error statis-

tics provides forward-looking methods to turn to. I do not claim that scientists will always be able to probe the needed errors successfully. My claim is that scientists do regularly tackle and often enough solve Duhemian problems, and that they do so by means of error statistical reasoning. Once we have set out the ingredients of an experimental framework (in chapter 5) we will see more clearly how an inquiry may be broken down so that each hypothesis is a local assertion about a particular error. There, and again in later chapters (e.g., chapters 6 and 13) we will return to Duhem's problem.

In the following subsection, I summarize the calculations that yield the results of Dorling's homework problem.

Calculations for the Homework Problem:

BACKGROUND ASSUMPTIONS:

Hypotheses A and T entail e, but e' is observed: $P(e' \mid A \text{ and } T) = 0$. A and T are statistically independent

ASSUMED PRIOR PROBABILITIES

$P(T) = .9$, $P(A) = .6$.

ASSUMED LIKELIHOODS:

a. $P(e' \mid A \text{ and } \sim T) = \varepsilon$ (very small number, e.g., .001).

b. i. $P(e' \mid \sim A \text{ and } T) = 50\varepsilon$

ii. $P(e' \mid \sim A \text{ and } \sim T) = 50\varepsilon$

Bayes's theorem:

$$P(T \mid e') = \frac{P(e' \mid T)\ P(T)}{P(e')}$$

From the above we get the following:

$$P(e') = P(e' \mid T)P(T) + P(e' \mid \sim T)P(\sim T).$$

$$\begin{aligned} P(e' \mid T) &= P(e' \mid A \text{ and } T)\ P(A) + P(e' \mid \sim A \text{ and } T)P(\sim A) \\ &= \quad 0 \quad\quad\quad + 50\varepsilon(.4) \\ &= 20\varepsilon. \end{aligned}$$

$$\begin{aligned} P(e' \mid \sim T) &= P(e' \mid A \text{ and } \sim T)P(A) + P(e' \mid \sim A \text{ and } \sim T)P(\sim A) \\ &= \varepsilon(.6) \quad\quad\quad + 50\varepsilon(.4) \\ &= 20.6\varepsilon. \end{aligned}$$

So

$$\begin{aligned} P(e') &= 20\varepsilon(.9) + 2.06\varepsilon \\ &= 20.06\varepsilon. \end{aligned}$$

The posterior probability of T can now be calculated:

$$P(e') = \frac{20\varepsilon(.9)}{20.06\varepsilon}$$
$$= 0.897.$$

Next we can calculate the posterior probability of A: By Bayes's theorem: $P(A \mid e') = \frac{P(e' \mid A)\ P(A)}{P(e')}$. Since

$$P(e' \mid A) = P(e' \mid A \text{ and } T)P(T) + P(e' \mid A \text{ and } \sim T)P(\sim T)$$
$$= \quad 0 \quad + \quad \varepsilon(.1)$$
$$= .1\varepsilon.$$

We get

$$P(A \mid e') = \frac{.06\varepsilon}{20.06\varepsilon}$$
$$= .003.$$

4.2 THOMAS KUHN MEETS THOMAS BAYES, INTRODUCTIONS BY WESLEY SALMON

I have thus far confined my criticism to the standard subjectivist Bayesian approach. There have been attempts to constrain the prior probabilities but with very limited success, especially when it comes to the Bayesian Way in philosophy of science. To discuss them here would require introducing technical ideas beyond the scope of our discussion. There is, however, one line of approach, developed by Wesley Salmon, that will tie together and illuminate a number of the themes I have taken up. My focus will be on his paper "Rationality and Objectivity in Science, *or* Tom Kuhn Meets Tom Bayes" (Salmon 1990).

As with the discussion in the previous section, Salmon's discussion is an attempt to employ the Bayesian Way to solve a philosophical problem, this time to answer Kuhn's challenge as to the existence of an empirical logic for science. Reflecting on the deep division between the logical empiricists and those who adopt the "historical approach," a division owing much to Kuhn's *Structure of Scientific Revolutions,* Salmon (1990) proposes "that a bridge could be built between the differing views of Kuhn and Hempel if Bayes's theorem were invoked to explicate the concept of scientific confirmation" (p. 175). The idea came home to Salmon, he tells us, during an American Philosophical Association (Eastern Division 1983) symposium on Carl Hempel, in which

Kuhn and Hempel shared the platform.[4] "At the time it seemed to me that this maneuver could remove a large part of the dispute between standard logical empiricism and the historical approach to philosophy of science" on the fundamental issue of confirmation (p. 175).

Granting that observation and experiment, together with hypothetico-deductive reasoning, fail adequately to account for theory choice, Salmon argues that the Bayesian Way can accommodate the additional factors Kuhn seems to think are required. In building his bridge, Salmon often refers to Kuhn's (1977) "Objectivity, Value Judgment, and Theory Choice." It is a fitting reference: in that paper Kuhn himself is trying to build bridges with the more traditional philosophy of science, aiming to thwart charges that he has rendered theory choice irrational.

Deliberately employing traditional terminology, Kuhn attempts to assuage his critics. He assures us that he agrees entirely that the standard criteria—accuracy, consistency, scope, simplicity, and fruitfulness—play a vital role in choosing between an established theory and a rival (p. 322). But as noted in chapter 2, Kuhn charges that these criteria underdetermine theory choice: they are imprecise, differently interpreted and differently weighed by different scientists. Taken together, they may contradict each other—one theory being most accurate, say, while another is most consistent with background knowledge. Hence theory appraisals may disagree even when agents ostensibly follow the same shared criteria. They function, Kuhn says, more like values than rules.

Here's where one leg of Salmon's bridge enters. The shared criteria of theory choice, Salmon proposes, can be cashed out, at least partly, in terms of prior probabilities. The conflicting appraisals that Kuhn might describe as resulting from different interpretations and weightings of the shared values, a Bayesian could describe as resulting from different assignments of prior probabilities. We have at least a partial bridge linking Bayes and Kuhn, but would a logical empiricist want to cross it?

Logical empiricists, it seems, would need to get around the Kuhnian position that the shared criteria are never sufficient to ground the choice between an accepted theory and a competitor, that consensus, if it occurs, always requires an appeal to idiosyncratic, personal factors beyond the shared ones. They would need to counter Kuhn's charge that in choosing between rival theories "scientists behave like philosophers," engaging in what I called "mere critical discourse" in chapter 2.

4. See "Symposium: The Philosophy of Carl G. Hempel," *Journal of Philosophy* 80, no. 10 (October 1983):555–72. Salmon's contribution is Salmon 1983.

Interestingly, Kuhn's single reference to a Bayesian approach is to combat criticism of his position. For the sake of argument, Kuhn says, suppose that scientists deploy some Bayesian algorithm to compute the posterior probabilities of rival theories on evidence and suppose that we could describe their choice between these theories as based on this Bayesian calculation (Kuhn 1977, 328). "Nevertheless," Kuhn holds that "the algorithms of individuals are all ultimately different by virtue of the subjective considerations with which each must complete the objective criteria before any computations can be done" (p. 329). So sharing Bayes's theorem does not count as a "shared algorithm" for Kuhn. Kuhn views his (logical empiricist) critic as arguing that since scientists often reach agreement in theory choice, the subjective elements are eventually eliminated from the decision process and the Bayesian posteriors converge to an objective choice. Such an argument, Kuhn says, is a non sequitur. In Kuhn's view, the variable priors lead different scientists to different theory choices, and agreement, if it does occur, results from sociopsychological factors, if not from unreasoned leaps of faith. Agreement, in other words, might just as well be taken as evidence of the further role of subjective and sociopsychological factors, rather than of their eventual elimination.

But perhaps building a logical empiricist bridge out of Bayesian bricks would not require solving this subjectivity problem. Perhaps Salmon's point is that by redescribing Kuhn's account in Bayesian terms, Kuhn's account need not be seen as denying science a logic based on empirical evidence. It can have a logic based on Bayes's theorem. It seems to me that much of the current appeal of the Bayesian Way reflects this kind of move: while allowing plenty of room for "extrascientific" factors, Bayes's theorem ensures at least some role for empirical evidence. It gives a formal model, we just saw, for reconstructing (after the fact) a given assignment of blame for an anomaly, and it may well allow for reconstructing Kuhnian theory choice. Putting aside for the moment whether a bridge from Bayes to Kuhn holds us above the water, let us see how far such a bridge would need to go.

Right away an important point of incongruity arises. While Kuhn talks of theory acceptance, the Bayesian talks only of probabilifying a theory—something Kuhn eschews. For the context of Kuhnian normal science, where problems are "solved" or not, this incongruity is too serious to remedy. But Salmon is talking about theory choice or theory preference, and here there seem to be ways of reconciling Bayes and Kuhn (provided radical incommensurabilities are put to one side), although Salmon does not say which he has in mind. One possibility would be to supplement the Bayesian posterior probability assessment

with rules for acceptance or preference (e.g., accept or prefer a theory if its posterior probability is sufficiently higher than that of its rivals).

A second possibility would be to utilize the full-blown Bayesian decision theory. Here, averaging probabilities and utilities allows calculating the average or expected utility of a decision. The Bayesian rule is to choose the action that *maximizes expected utility.* Choosing a theory would then be represented in Bayesian terms as adopting the theory that the agent feels maximizes expected utility. If it is remembered that, according to Kuhn, choosing a theory means deciding to work within its paradigm, this second possibility seems more apt than the first. The utility calculation would provide a convenient place to locate the variety of values—those shared as well as those of "individual personality and biography"—that Kuhn sees as the basis for theory choice.

Even this way of embedding Kuhn in a Bayesian model would not quite reach the position Kuhn holds. In alluding to the Bayesian model, Kuhn (1977) concedes that he is tempering his position somewhat, putting to one side the problems of radically theory-laden evidence and incommensurability. Strictly speaking, comparing the expected utilities of choosing between theories describes a kind of comparison that Kuhn deems impossible for choosing between incommensurables. It is doubtful that a genuine Kuhnian conversion is captured as the result of a Bayesian conditionalization. Still, the reality of radical incommensurability has hardly been demonstrated. So let us grant that the subjective Bayesian Way, with the addition of some rule of acceptance such as that offered by Bayesian decision theory, affords a fairly good bridge between Bayes and a slightly-tempered Kuhn. Note also that the Kuhnian problems of subjectivity and relativism are rather well modeled—though not solved—by the corresponding Bayesian problems. The charge that Kuhn is unable to account for how scientists adjudicate disputes and often reach consensus seems analogous to the charge we put to the subjective Bayesian position. (For a good discussion linking Kuhn and Bayes, see Earman 1992, 192–93.)

But this is not Salmon's bridge. Our bridge pretty much reaches Kuhn, but the toll it exacts from the logical empiricist agenda seems too dear for philosophers of that school to want to cross it. Salmon's bridge is intended to be free of the kinds of personal interests that Kuhn allows, and as such it does not go as far as reaching Kuhn's philosophy of science. But that is not a mark against Salmon's approach, quite the opposite. A bridge that really winds up in Kuhnian territory is a bridge too far: a utility calculation opens theory choice to all manner of interests and practical values. It seems the last thing that would appeal to those wishing to retain the core of a logical empiricist philos-

ophy. (It opens too wide a corridor for the enemy!) So let us look at Salmon's bridge as a possible link, not between a tempered Kuhn and Bayes, but between logical empiricism and a tempered Bayesianism. Before the last brick is in place, I shall question whether the bridge does not actually bypass Bayesianism altogether.

4.3 SALMON'S COMPARATIVE APPROACH AND A BAYESIAN BYPASS

Salmon endorses the Kuhnian position that theory choice, particularly among mature sciences, is always a matter of choosing between rivals. Kuhn's reason, however, is that he regards rejecting a theory or paradigm in which one had been working without accepting a replacement as tantamount to dropping out of science. Salmon's reason is that using Bayes's theorem comparatively helps cancel out what he takes to be the most troubling probability: the probability of the evidence e given not-T ("the catchall"). (Salmon, like me, prefers the term hypotheses to theories, but uses T in this discussion because Kuhn does. I shall follow Salmon in allowing either to be used.)

Because of some misinterpretations that will take center stage later, let us be clear here on the probability of evidence e on the catchall hypothesis.[5] Evidence e describes some outcome or information, and not-T, the catchall, refers to the disjunction of all possible hypotheses other than T, including those not even thought of, that might predict or be relevant to e. This probability is not generally meaningful for a frequentist, but is necessary for Bayes's theorem.[6] Let us call it the *Bayesian catchall factor* (with evidence e):[7]

> Define the *Bayesian catchall factor* (in assessing T with evidence e) as
>
> $P(e \mid \text{not-}T)$.

Salmon, a frequentist at heart, rejects the use of the Bayesian catchall factor.

> What is the likelihood of any given piece of evidence with respect to the catchall? This question strikes me as utterly intractable; to answer it we would have to predict the future course of the history of science. (Salmon 1991, 329)

5. To my knowledge, it was L. J. Savage who originated the term *catchall*.
6. See chapter 6.
7. I take this term from that of the Bayes factor, which is the ratio of the Bayesian catchall factor and $P(e \mid T)$.

This recognition is a credit to Salmon, but since the Bayesian catchall factor is vital to the general Bayesian calculation of a posterior probability, his rejecting it seems almost a renunciation of the Bayesian Way. The central role of the Bayesian catchall factor is brought out in writing Bayes's theorem as follows:

$$P(T \mid e) = \frac{P(e \mid T)\ P(T)}{P(e \mid T)\ P(T) + \mathbf{P(e \mid not\text{-}T)}\ P(\text{not-}T).}$$

Clearly, the lower the value of the Bayesian catchall factor, the higher the posterior probability in T, because the lower its value, the less the denominator in Bayes's theorem exceeds the numerator. The subjectivist "solution" to Duhem turned on the agent assigning a very small value to the Bayesian catchall factor (where the evidence was the anomalous result e'), because that allowed the posterior of T to remain high despite the anomaly. Subjective Bayesians accept, as a justification for this probability assignment, that agents believe there to be no plausible rival to T that they feel would make them expect the anomaly e'. This is not good enough for Salmon.

In order to get around such a subjective assignment (and avoid needing to predict the future course of science), Salmon says we should restrict the Bayesian Way to looking at the ratio of the posteriors of two theories T_1 and T_2: In the ratio of the posteriors of the two theories, we get a canceling out of the Bayesian catchall factors (the probability of e on the catchall).[8] Let us see what the resulting comparative assessment looks like. Since the aim is no longer to bridge Kuhn, we can follow Salmon in talking freely about either theories or hypotheses. Salmon's Bayesian algorithm for theory preference is as follows (to keep things streamlined, I drop the explicit statement of the background variable B):

Salmon's Bayesian algorithm for theory preference (1990, 192):

Prefer T_1 to T_2 whenever $P(T_1 \mid e)/P(T_2 \mid e)$ exceeds 1, where:

8. This is because

$$P(T_i \mid e) = \frac{P(e \mid T_i)P(T_i)}{P(e \mid T_i)P(T_i) + P(e \mid \sim T_i)P(\sim T_i).}$$

Note that the denominator equals $P(e)$. Since that is so for the posterior of T_1 as well as for T_2, the result of calculating the ratio is to cancel $P(e)$, and thereby cancel the probabilities of e on the catchalls.

$$\frac{P(T_1 \mid e)}{P(T_2 \mid e)} = \frac{P(T_1)\ P(e \mid T_1)}{P(T_2)\ P(e \mid T_2)}.$$

To start with the simplest case, suppose that both theories T_1 and T_2 entail e.[9] Then $P(e \mid T_1)$ and $P(e \mid T_2)$ are both 1. These two probabilities are the *likelihoods* of T_1 and T_2, respectively.[10] Salmon's rule for this special case becomes:

> *Special case:* Salmon's rule for relative preference (where each of T_1 and T_2 entails e):
>
> Prefer T_1 to T_2 whenever $P(T_1)$ exceeds $P(T_2)$.

Thus, in this special case, the relative preference is unchanged by evidence e. You prefer T_1 to T_2 just in case your prior probability in T_1 exceeds that of T_2 (or vice versa). Note that this is a general Bayesian result that we will want to come back to. In neutral terms, it says that if evidence is entailed by two hypotheses, then *that evidence* cannot speak any more for one hypothesis than another—according to the Bayesian algorithm.[11] If their appraisal differs, it must be due to some difference in prior probability assignments to the hypotheses. This will not be true on the error statistics model.

To return to Salmon's analysis, he proposes that where theories do not entail the evidence, the agent consider auxiliary hypotheses (A_1 and A_2) that, when coinjoined with each theory (T_1 and T_2, respectively), would entail the evidence. That is, the conjunction of T_1 and A_1 entails e, and the conjunction of T_2 and A_2 entails e. This allows, once again, the needed likelihoods to equal 1, and so to drop out. The relative appraisal of T_1 and T_2 then equals the ratio of the prior probabilities of the conjunctions of T_1 and A_1, and T_2 and A_2. We are to prefer that conjunction (of theory and auxiliary) that has the higher prior probability.[12] In short, in Salmon's comparative analysis the weight is taken from the likelihoods and placed on the priors, making the appraisal even more dependent upon the priors than the noncomparative Bayesian approach.

9. While this case is very special, Salmon proposes that it be made the standard case by conjoining suitable auxiliaries to the hypotheses. I will come back to this in a moment.

10. Note that likelihoods of hypotheses are *not* probabilities. For example, the sum of the likelihoods of a set of mutually exclusive, collectively exhaustive hypotheses need not equal 1.

11. This follows from the likelihood principle to be discussed in later chapters.

12. For simplicity, we could just replace T_1, T_2, in the statement of the special case, with the corresponding conjunctions T_1 and A_1, and T_2 and A_2, respectively.

Problems with the Comparative Bayesian Approach

Bayesians will have their own problems with such a comparative Bayesian approach. How, asks Earman (1992, 172), can we plug in probabilities to perform the usual Bayesian decision theory? But Earman is reluctant to throw stones, confessing that "as a fallen Bayesian, I am in no position to chide others for acts of apostasy" (p. 171). Earman, with good reason, thinks that Salmon has brought himself to the brink of renouncing the Bayesian Way. Pursuing Salmon's view a bit further will show that he may be relieved of the yoke altogether.

For my part, the main problem with the comparative approach is that we cannot apply it until we have accumulated sufficient knowledge, by some non-Bayesian means, to arrive at the prior probability assignments (whether to theories or theories conjoined with auxiliaries). Why by some non-Bayesian means? Couldn't prior probability assessments of theories and auxiliaries themselves be the result of applying Bayes's theorem? They could, but only by requiring a *re*introduction of the corresponding assignments to the Bayesian catchall factors—the very thing Salmon is at pains to avoid. The problems of adjudicating conflicting assessments, predicting the future of science, and so on, remain.

Could not the prior probability assignments be attained by some more hard-nosed assessment? Here is where Salmon's view becomes most interesting. While he grants that assessments of prior probabilities, or, as he prefers, plausibilities, are going to be relative to agents, Salmon demands that the priors be constrained to reflect objective considerations.

> The frightening thing about pure unadulterated personalism is that nothing prevents prior probabilities (and other probabilities as well) from being determined by all sorts of idiosyncratic and objectively irrelevant considerations (Salmon 1990, 183)

such as the agent's mood, political disagreements with or prejudices toward the scientists who first advanced the hypothesis, and so on.

> What we want to demand is that the investigator make every effort to bring all of his or her *relevant* experience in evaluating hypotheses to bear on the question of whether the hypothesis under consideration is of a type likely to succeed, and to leave aside emotional irrelevancies. (P. 183)

Ever the frequentist, Salmon proposes that prior probabilities "can be understood as our best estimates of the frequencies with which cer-

tain kinds of hypotheses succeed" (p. 187).[13] They may be seen as personalistic so long as the agent is guided by "the aim of bringing to bear all his or her experience that is relevant to the success or failure of hypotheses similar to that being considered." According to Salmon, "On the basis of their training and experience, scientists are qualified to make such judgments" (p. 182).

But are they? How are we to understand the probability Salmon is after? The context may be seen as a single-universe one. The members of this universe are hypotheses similar to the hypothesis *H* being considered, presumably from the population of existing hypotheses. To assign the prior probability to hypothesis *H*, I imagine one asks oneself, What proportion of the hypotheses in this population are (or have been) successful? Assuming that *H* is a random sample from the universe of hypotheses similar to *H*, this proportion equals the probability of interest. Similar to hypothesis *H* in what respects? Successful in what ways? For how long? The reference class problem becomes acute.

I admit that this attempt at a frequentist prior (also found in Hans Reichenbach) has a strong appeal. My hunch is that its appeal stems from unconsciously equating this frequency with an entirely different one, and it is this different one that is really of interest and, at the same time, is really obtainable.

Let us imagine that one had an answer to Salmon's question: what is the relative frequency with which hypotheses relevantly similar to *H* are successful (in some sense)? Say the answer is that 60 percent of them are. If *H* can be seen as a fair sample from this population, we could assign *H* a probability of .6. Would it be of much help to know this? I do not see how. I want to know how often *this* hypothesis will succeed.

What might an error statistician mean by the probability that this hypothesis *H* will succeed? As always, for a frequentist, probability applies to outcomes of a type of experiment. (They are sometimes called "generic" outcomes.) The universe or population here consists of possible or hypothetical experiments, each involving an application of hypothesis *H*. Success is some characteristic of experimental outcomes. For example, if *H* is a hypothesized value of a parameter μ, a successful outcome might be an outcome that is within a specified margin of error of μ. The probability of success construed this way is just the ordinary

13. This was also Reichenbach's view. He did not consider that enough was known at present to calculate such a probability, but thought that it might be achievable in the future.

probability of the occurrence of certain experimental outcomes. (Further discussion occurs in chapter 5.) Indeed, for the error theorist, the only kinds of things to which probabilities apply are things that can be modeled as experimental outcomes. Knowledge of *H*'s probable success is knowledge of the probability distributions associated with applying *H* in specific types of experiments. Such knowledge captures the spirit of what C. S. Peirce would call the "experimental purport" of hypothesis *H*.

Two Meanings of the Probability That a Hypothesis Is Successful. Let us have a picture of our two probabilities. Both can be represented as one-urn models. In Salmon's urn are the members of the population of hypotheses similar to *H*. These hypotheses are to be characterized as successful or not, in some way that would need to be specified. The probability of interest concerns the relative frequency with which hypotheses in this urn are successful. This number is taken as the probability that *H* is successful.

In my urn are members of a population of outcomes (a sample space) of an experiment. Each outcome is defined as successful or not according to whether it is close to what *H* predicts relative to a certain experiment (for simplicity, omit degrees of closeness). The probability of interest concerns the relative frequency with which outcomes in this urn are successful. Hypothesis *H* can be construed as asserting about this population of outcomes that with high probability they will be successful (e.g., specifiably close to what *H* predicts). The logic of standard statistical inference can be pictured as selecting one outcome, randomly, from this "urn of outcomes" and using it to learn whether what *H* asserts is correct.

Take one kind of hypothesis already discussed, that a given effect is real or systematic and not artifactual. In particular, take Hacking's hypothesis *H* (discussed in chapter 3).

H: dense bodies are real structures in blood cells, not artifacts. We have no idea what proportion of hypotheses similar to *H* are true, nor do we have a clue as to how to find out, nor what we would do if we did. In actuality, our interest is not in a probabilistic assignment to *H*, but in whether *H* is the case. We need not have infallible knowledge about *H* to learn about the correctness of *H*.

We ask: what does the correctness of *H* say about certain experimental results, ideally, those we can investigate? One thing Hacking's *H* says is that dense bodies will be detected even using radically different physical techniques, or at least that they will be detected with high reliability. Experimenting on dense bodies, in other words, will *not* be

use the latter frequentist notion in applying Bayes's theorem. Nevertheless, it may be used in Salmon's comparative approach (where the likelihoods drop out), and doing so yields a very natural bridge connecting his approach to that of error statistics.

To see in a simple way what this natural bridge looks like, let the two hypotheses H_1 and H_2 entail evidence e (it would be adequate to have them merely fit e to some degree). Then, on Salmon's comparative Bayesian approach, H_1 is to be preferred to H_2 just in case the prior probability assessment of H_1 exceeds H_2. The assignment of the prior probabilities must not contain irrelevant subjective factors, says Salmon, but must be restricted to assessing whether the hypotheses are likely to be successful. Hypothesis H_1 is to be preferred to H_2 just in case H_1 is accorded a higher probability of success than H_2. Now let us substitute my error statistical construal of probable success (for some specified measure of "successful outcome"). Evaluating H's probable success (or H's reliability) means evaluating the relative frequency with which applications of H would yield results in accordance with (i.e., specifiably close to) what H asserts. As complex as this task sounds, it is just the kind of information afforded by experimental knowledge of H. The task one commonly sets for oneself is far less technically put. The task, informally, is to consider the extent to which specific obstacles to H's success have been ruled out. Here is where the kind of background knowledge I think Salmon has in mind enters. What training and experience give the experimenter is knowledge of the specific ways in which hypotheses can be in error, and knowledge of whether the evidence is so far sufficient to rule out those errors.

To put my point another way, Salmon's comparative approach requires only the two prior probabilities or plausibilities to be considered, effectively wiping out the rest of the Bayesian calculation. The focus is on ways of assessing the plausibilities of the hypotheses or theories themselves. However, Salmon's approach gives no specific directions for evaluating the plausibilities or probable success of the hypotheses. Interpreting probable success in the way I recommend allows one to work out these directions. Salmon's comparative appraisal of H_1 against a rival H_2 would become: prefer H_1 to H_2 just to the extent that the evidence gives a better indication of H_1's likely success than H_2's.

Further, the kinds of evidence and arguments relevant to judge H's success, in my sense, seem quite congenial to what Salmon suggests should go into a plausibility assessment. In one example Salmon refers explicitly to the way in which standard (non-Bayesian) significance tests may be used to give plausibility to hypotheses (Salmon 1990, 182). In particular, a statistically significant association between sac-

charin and bladder cancer in rats, he says, lends plausibility to the hypothesis H that saccharin in diet drinks increases the risk of bladder cancer in humans. Provided that errors of extrapolating from rats to humans and from high to low doses are satisfactorily ruled out, a statistically significant association may well provide evidence that H has been shown, *that H will be successful in our sense*. This success may be cashed out in different ways, because the truth of H has different implications. One implication of the correctness of H here is that were populations to consume such and such amount of saccharin the incidence of bladder cancer would be higher than if they did not. My point is that such experiments are evidence for the correctness of H in this sense. Such experiments do not provide the number Salmon claims to be after, the probability that hypotheses similar to the saccharin hypothesis are successful. So even if that probability were wanted (I claim it is not), that is not what the saccharin experiments provide.[14]

By allowing for this error statistical gloss of "H's probable success," the reader should not be misled into viewing our account as aiming to assign some quantitative measure to hypotheses—the reverse is true. My task here was to erect a bridge between an approach like Salmon's and the testing account I call error statistics. By demonstrating that the

14. In the case of the saccharin hypothesis, it might look as if Salmon's frequentist probability is obtained. That, I think, is because of a tendency to slide from one kind of probability statement to another. Consider hypotheses of the form x causes cancer in humans. They are all similar to H: saccharin causes cancer in humans. But what should be included in the reference set for getting Salmon's probability? Might x be anything at all? If so, then only a very tiny proportion would be successful hypotheses. That would not help in assessing the plausibility of H. I suggest that the only way this probability makes sense is if hypotheses "similar to H" refers to hypotheses similarly grounded or tested. In trying to specify the reference set in the case of the saccharin hypothesis we might restrict it to those causal hypotheses (of the required form) that have been shown to hold about as well as H. So it would consist of causal claims where a statistically significant correlation is found in various animal species, where certain dosage levels are used, where certain extrapolation models are applied (to go from animal doses to human doses as well as from rats to humans), where various other errors in identifying carcinogens are ruled out, and so on. Notice how these lead to a severity assessment.

A relative frequency question of interest that can be answered, at least qualitatively, is this: What is the relative frequency with which hypotheses of this sort (x causes cancer in humans) pass experimental tests $E_1, \ldots, E_n$ as well as H does, and yet do not succeed (turn out to be incorrect)? One minus this gives the severity of the test H passes. The question boils down to asking after the severity of the test H passes (where, as is common, several separate tests are taken together).

It does not matter that the hypotheses here differ. Error probabilities of procedures hold for applications to the same or *different* hypotheses. Neyman (e.g., 1977) often discussed the mistake in thinking they hold only for the former.

role Salmon gives to plausibility assessments is better accomplished by an assessment of the reliability of the tests hypotheses pass, I mean to show that the latter is all that is needed.

There are plenty of advantages to the testing account of scientific inference. First, by leading to accepting hypotheses as approximately correct, as well indicated, or as likely to be successful—rather than trying to assign some quantity of support or probability to hypotheses—it accords with the way scientists (and the rest of us) talk. Second, reporting the quality of the tests performed provides a way of communicating the evidence (summarizing the status of the problem to date) that is intersubjectively testable. A researcher might say, for example, that the saccharin rat study gives good grounds for holding that there is a causal connection with cancer in rats, but deny that the corresponding hypothesis about humans has been severely tested. This indicates what further errors would need to be ruled out (e.g., certain dose-response models are wrong).

Now it is open to a Bayesian to claim that the kinds of arguments and evidence that I might say give excellent grounds for the correctness of H, for accepting H, or for considering H to have passed a severe test can be taken as warranting a high prior probability assignment in H. For example, "there are excellent grounds for H" may be construed as "H has high prior probability" (say, around .9). (That Bayesians implicitly do this in their retrospective reconstructions of episodes is what gives their prior probability assessments their reasonableness.) Used in a purely comparative approach such as Salmon's, it might do no harm. However, there is nothing Bayesian left in this comparative approach! It is, instead, a quantitative sum-up of the quality of *non-Bayesian* tests passed by one hypothesis compared with those passed by another. (Whether such a non-Bayesian assessment of Bayesian priors could even be made to obey the probability calculus is not clear.)

To call such an approach Bayesian, even restricting it to comparisons, would be misleading. It is not just that the quantitative sum-up of H's warrant is not arrived at via Bayes's theorem. It is, as critics of error statistics are happy to demonstrate, that the principles of testing used in the non-Bayesian methods conflict with Bayesian principles.[15] (I will have much more to say about this later, e.g., in chapter 10.) The Bayesian Way supposes, for any hypothesis one wishes to consider,

15. To anticipate a little, the Bayesian Way follows the likelihood principle, which conflicts with error-probability principles. Quoting Savage: "Practically none of the 'nice properties' respect the likelihood principle" (1964, 184). The "nice properties" refer to error characteristics of standard statistical procedures, such as unbiasedness and significance levels. I return to this in chapter 10.

that a Bayesian prior is available for an agent, and that an inference can be made. In general, however, there are not going to be sufficient (non-Bayesian) grounds to assign even a rough number to such hypotheses. We are back to the problem of making it too difficult to get started when, as is commonly the case, one needs a forward-looking method to begin learning something.

Bayesian Heretics, Fallen and Disgruntled Bayesians

The Bayesian landscape is littered with Bayesians who variably describe themselves or are described by others as fallen, heretical, tempered, nonstrict, or whatnot. Many Bayesians in this category came to the Bayesian Way in the movements led by Carnap and Reichenbach. Assigning probabilities to hypotheses was a natural way of avoiding the rigidities of a hypothetico-deductive approach. Inadequacies in the two main objective ways philosophers tried to define the prior probabilities—Carnapian logical or Reichenbachian frequentist—have left some in limbo: wanting to avoid the excesses of personalism and not sure how non-Bayesian statistics can help. Those Bayesians do not see themselves as falling under the subjectivist position that I criticized earlier. I invite them to try out the natural bridge proffered above, to see where it may lead.

What is really being linked by this bridge? Might it be said to link the cornerstone of logical empiricism on the one hand and the centerpiece of the New Experimentalism on the other? Such a bridge, as I see it, would link the (logical empiricist) view that the key to solving problems in philosophy of science is an inductive-statistical account of hypothesis appraisal with the view that the key to solving problems in philosophy of science is an understanding of the nature and role of experiment in scientific practice. It provides a way to *model* Kuhn's view of science—*where he is correct*—as well as a way to *"solve* Kuhn" where he challenges the objectivity and rationality of science.

In this chapter and the last I have brought out the main shortcomings of appeals to the Bayesian Way in modeling scientific inference and in solving problems about evidence and inference. Understanding these shortcomings also puts us in a better position to see what would be required of any theory of statistics that purports to take a leading role in an adequate philosophy of experiment. For one thing, we need an account that explicitly incorporates the intermediate theories of data, instruments, and experiment that are required to obtain experimental evidence in the first place. For another, the account must enable us to address the question of whether auxiliary hypotheses or experimental assumptions are responsible for observed anomalies from a

hypothesis H, quite apart from how credible we regard hypothesis H. In other words, we need to be able to split off from some primary inquiry or test those questions about how well run the experiment was, or how well its assumptions were satisfied. Let us now turn to an experimental framework that will lend itself to these requirements.