REASON, JUDGEMENT AND BAYES'S LAW*

HAROLD I. BROWN†‡

Department of Philosophy
Northern Illinois University

This paper argues that when used judiciously Bayes's law has a role to play in the evaluation of scientific hypotheses. Several examples are presented in which a rational response to evidence requires a judgement whether to apply Bayes's law or whether, for example, to redistribute prior probabilities. The paper concludes that reflection on Bayes's law illustrates how an adequate account of the rational evaluation of hypotheses requires an account of judgement—a point which several philosophers have noted despite few attempts to develop an adequate theory of judgement.

1. Preliminaries. A traditional and still vigorous program in the philosophy of science seeks to capture the rational evaluation of scientific hypotheses in a formalism that will allow us to assess the impact of evidence on hypotheses. Bayesian confirmation theory is one avenue through which this program is currently being pursued. While many forms of Bayesianism have been developed (Good 1983), a common thread appears in the case of confirmation theory: the central role played by Bayes's law in evaluating the probability of hypotheses. Bayes's law is a theorem of the probability calculus, which can be written as follows:

$$P(h|e) = P(h)P(e|h)/P(e), \text{ provided } P(e) \neq 0. \qquad (1)$$

In this equation the hypothesis h is being evaluated on the basis of the evidence statement e. To use this formalism we begin with an estimate of the probability that our hypothesis is true independently of our knowledge of e. This probability, $P(h)$, is the "prior probability of h"; $P(e|h)$, the "likelihood", is the probability that e is true conditional on h; $P(e)$ is an estimate of the probability that e is true. Now $P(h|e)$ is our new estimate of the probability that h is true once e has been taken into account. Thus Bayes's law provides an algorithm for adjusting the probability of h as evidence accumulates. Each of the probabilities in this equation is to be evaluated in the light of accepted background knowledge. Including the background knowledge in (1) yields $P(h|e \,\&\, k) = P(h/k)P(e|h \,\&\, k)/$

*Received April 1993; revised November 1993.
†I want to thank C. A. Hooker and an anonymous *Philosophy of Science* referee for helpful comments on earlier drafts of this paper.
‡Send reprint requests to the author, Department of Philosophy, Northern Illinois University, DeKalb, IL 60115, USA.

Philosophy of Science, 61 (1994) pp. 351–369

$P(e/k)$. I will omit k in order to simplify the formula. This will not affect any of the points made in this paper.

As is the case for any algorithm, application of (1) requires that the input values be determined before the algorithm is applied. This suggests that our value for $P(h|e)$ will be no more reliable than the least reliable of our inputs. Thus an examination of the epistemic status of these inputs is a crucial issue in evaluating the Bayesian program. Indeed, the status of the inputs provides one major source of disagreement among Bayesians and I can specify the version of Bayesian confirmation theory that I will be concerned with in this paper by considering the status of these inputs. As a first step note that Salmon has recently characterized one approach as "the standard Bayesian approach in which all of the probabilities are taken to be personal probabilities" (1990, 181). However, recent literature presents a widely defended form of Bayesianism that differs from Salmon's characterization in one respect: A substantial number of prominent Bayesians deny that the likelihoods are personal probabilities. To see why, let us distinguish two cases: that in which h is a deterministic hypothesis and that in which h is a statistical hypothesis.

In typical examples of the former case, either h entails e and $P(e|h)$ is one, or h entails $-e$ and $P(e|h)$ is zero. Thus the value of $P(e|h)$ is not subjective in this case because this value is assessed on the basis of a deduction. When h is a statistical hypothesis, h does not entail e but rather a probability value for e. For example, if our evidence consists of two flips of a coin yielding two heads, and h is the hypothesis that this is a fair coin, we can calculate that $P(e|h) = 0.25$. Again, the value of $P(e|h)$ is assessed on the basis of a deductive argument. The view that the likelihoods are not subjective has been defended by Howson and Urbach (1989, 119–120), who consider the ability to provide a unified treatment of deterministic and statistical hypotheses to be one of the advantages of the Bayesian approach. Howson has recently reiterated the point, "The only endogenously determined quantities in the Personalistic Bayesian theory are the likelihood terms $P(e/h)$ . . . and the probabilities of necessary truths and falsehoods relative to the individual's background information" (1990, 228). This approach is also both defended and criticized in Earman (1992). Indeed, Earman introduces the thesis that likelihoods are also personal probabilities as a *criticism* of the view he discusses (ibid., 140–141).

Until section 5 of this paper I will use the term "Bayesian" for a view which deviates from Salmon's "standard approach" in one respect: The view holds that all of the inputs to Bayes's law *except* the likelihoods are subjective. Let us look briefly at some of the issues raised by the subjective status of the remaining inputs into Bayes's law.

Much discussion of Bayesianism concerns the possibility of overcoming the subjectivity of these inputs as evidence accumulates. It will be useful to develop our discussion of these inputs in terms of a familiar distinction from traditional confirmation theory: that between the context of justification and the context of discovery. The former context is limited to the logical assessment of hypotheses against the evidence; according to Bayesians this assessment occurs through applications of Bayes's law. Originally the context of discovery was limited to the processes by which scientists arrive at hypotheses. The main point of the distinction lay in the claim that this is an arational process; the rationality of our epistemic attitudes toward hypotheses depended solely on the evidence and the assessment carried out in the context of justification. This distinction is central to the Bayesian view, but Bayesians extend the set of items located in the context of discovery to include subjective estimates of the remaining inputs into the Bayesian formula and, as we will see as we proceed, other items as well. If a Bayesian account of the rationality of science is to succeed, Bayesians must show that accumulating evidence and the application of Bayes's law are sufficient to overcome the subjectivity of these inputs. Let us consider, then, the two remaining inputs to the Bayesian algorithm.

The most controversial (and most heavily discussed) of the required inputs is $P(h)$: the estimate of the probability of h prior to consideration of the evidence e. The value of $P(h)$ is estimated by the individual evaluator on the basis of available beliefs and any other considerations deemed relevant. The subjective nature of this estimate might seem to guarantee a subjective output from the algorithm, but a standard Bayesian response to this objection is available: We only estimate $P(h)$ once for a given hypothesis and that estimate becomes less significant as evidence accumulates. Given each new item of evidence, we apply Bayes's law again using the most recently calculated value of $P(h|e)$ as the current value of $P(h)$. Evidence, it is argued, eventually "swamps" the initial estimate of $P(h)$ so that all investigators converge on the same value for $P(h)$ whatever their initial estimates. Indeed, several theorems show that convergence must occur under certain conditions. The best known theorem is due to Savage (1954, 46–50); this theorem as well as more recent convergence results are discussed and their significance assessed in Earman (1992, 141–149, *et passim*). In general, Bayesians tend to rest content with the swamping thesis (e.g., Howson and Urbach 1989, 235–236) while critics question the relevance of these theorems to real scientific inference. Without adding to this literature, I will simply note a perspective on the swamping thesis that will be important as we proceed.

Those who rely on the swamping thesis share a central strategic move with Popper. On Popper's view, scientific hypotheses are proposed nonrationally and then evaluated by deducing testable consequences from them.

False hypotheses are eliminated when we deduce a conclusion that we can show to be false. Thus we need not worry about the nonrational source of our hypotheses since false hypotheses will be overwhelmed by the evidence. Now Popper sought to do away with induction and capture the logic of scientific justification wholly in deduction. Bayesians typically view themselves as providing an inductive logic, but deduction is the only logic required for a Bayesian "induction". Deductive inferences occur when we apply Bayes's law and when we determine the value of $P(e|h)$ by deducing either e or the probability of e from h. The remaining inputs to an application of the Bayesian algorithm are accepted nonrationally according to Bayesians' own standards. (I develop this point for $P(e)$ shortly.) In other words, on both the Popperian and Bayesian views deduction is the only logic required in the context of justification. Thus the key epistemological question becomes the status of the premises of these deductions; after all, any conclusion can be validly deduced from some premises. Both approaches attempt to solve this problem by arguing that incorrect inputs from the context of discovery will eventually be eliminated by the testing process. Howson (1990, 224, 243) noted some of these parallels with Popper.

The remaining input into equation (1) is $P(e)$ and its status is also controversial. Much of this controversy concerns the case in which we set $P(e)$ equal to one. Some argue that once we have carried out the relevant experiments and observations and decided that e is true, we should set $P(e)$ equal to one. However, this has the disadvantage that setting $P(e)$ equal to one in equation (1) is equivalent to dropping the term altogether. Moreover, as Glymour (1980, 86) noted when we evaluate a deterministic hypothesis that successfully predicts e, $P(e|h)$ is also one and we get no adjustment to the prior probability as a result of applying Bayes's law.

One Bayesian response, advocated by Garber (1983) and others, is that even if h entails e, we need not know that this is the case and thus we need not set $P(e|h)$ equal to one. While such logical learning does occur, however, many important scientific cases exist in which e has in fact been deduced from h so that we do know that $P(e|h) = 1$. A more common Bayesian view is that we must assess $P(e)$ on the basis of our background beliefs, abstracting from any knowledge of whether e is in fact true or false. (See, e.g., Howson and Urbach 1989, 270–275; and Howson 1991—which includes a reply to Garber. Eells 1990 provides replies both to Garber and to Jeffrey's 1983 version of this approach. See also Glymour 1980, 91–93, and Niiniluoto 1983. Earman 1992, chap. 5, provides a comprehensive discussion of these options.) I will raise some different issues in this paper. For present purposes two points are important. First, the evaluation of $P(e)$ constitutes a second place at which a subjective evaluation of a probability is required before we can apply the Bayesian algorithm. Bayesians have not applied the swamping approach to $P(e)$; I will return to $P(e)$ in section 3. Second, even when e is known to be true, Bayesians deny that we should substitute one for $P(e)$ in Bayes's law; this point will be especially important in section 4. (Another issue concerns the fact that we get greater confirmation of h for lower values of $P(e)$, ceteris paribus. See Howson and Urbach 1989, 86–88, for a Bayesian response. I return to this topic at the end of sec. 3.)

With this background, I can now state the *main thesis* of this paper. I will argue that while Bayes's law may have an important role to play in the rational evaluation of scientific hypotheses, use of this law can provide only a small part of such rational evaluation. We will see that applications of Bayes's law to accumulating evidence are not ipso facto rational. Indeed, situations arise in which serious questions must be addressed as to whether applying Bayes's law is reasonable and these questions cannot be resolved by further applications of the law. I begin my argument for this thesis with another look at the prior probabilities.

2. Hypotheses and Prior Probabilities. On the Bayesian view, hypothesis testing amounts to a competition between a set of explicitly formulated hypotheses. In one respect this is true for any confirmation theory: We cannot evaluate an hypothesis that has not been formulated. (We can include a catchall hypothesis to cover the possibility that all of the hypotheses formulated thus far may be wrong. But inclusion of the catchall will not affect the issues to be discussed in the present section.) Nevertheless, on a Bayesian approach the set of hypotheses under explicit consideration plays an especially important role because the sum of the prior probabilities assigned to these hypotheses must not exceed one. Thus we cannot assign prior probabilities on a case-by-case basis. Rather, we must make these assignments holistically, reflecting on the complete set of hypotheses under consideration and on our assessment of their relative probabilities. In effect, when we assign prior probabilities we are distributing a limited pool of available probability values. As a result, any decision to entertain a new hypothesis will often require more than just assigning another prior probability value; it may require some redistribution of prior probabilities previously assigned (see Earman 1992, 195–198, for particularly striking examples). In other words, on the Bayesian view, the decision to entertain a new hypothesis, or to revive a previously rejected hypothesis, will require a more complex set of decisions in the context of discovery than just deciding to entertain another hypothesis. Let us explore some variations on this situation, beginning with the case in which we decide to reconsider a previously rejected hypothesis.

Bayes's law can be applied only to hypotheses that have some prior probability greater than zero. This is clear from equation (1) where setting

$P(h) = 0$ ends discussion of that hypothesis. Savage describes this as "an example of the general principle that, if some event is regarded as virtually impossible, then no evidence whatsoever can lend it credibility" (1954, 47). Savage describes this as a "common-sense principle", and the point is reiterated a few years later in a joint paper, "roughly, once something is regarded as impossible, no evidence can reinstate its credibility" (Edwards et al. 1963, 218).

Yet many important scientific situations arise in which observational evidence should lead us to reconsider an hypothesis that we once viewed as having no possibility of being correct. For example, at one time many Aristotelians would have accorded a probability of zero to the suggestion that a rock dropped from the top of the mast of a moving ship would land at the foot of the mast. According to Aristotelian physics, a stone landing at the foot of the mast would have been simultaneously engaged in natural and violent motion. Since these two forms of motion are, by definition, contraries, it is *logically impossible* that they occur in a moving object at the same time. Still, a person who initially held an Aristotelian view and then observed the experiment would do well to reconsider *because of* the outcome. Moreover, once the point has been made, other examples leap quickly to the mind. Before Galileo's telescopic observations many astronomers would have accorded a probability of zero to the hypotheses that there are mountains on the moon or spots on the sun. Yet some of these astronomers changed their minds in response to the telescopic observations. Similarly, a late nineteenth-century physicist might well have rejected the claim that mass is a function of velocity, that space and time are aspects of a more fundamental feature of the physical world, or that some events are uncaused, but reconsidered this rejection on the basis of observational evidence.

The above examples bring out one way in which the Bayesian view of scientific rationality is too narrow. None of these decisions to reconsider a previously rejected hypothesis are rational if the rational evaluation of scientific hypotheses *requires* an application of Bayes's law. Nor, on Bayesian grounds, could there be any failure of rationality in the refusal to accord any probability at all to the claim that, say, the sun and moon are imperfect—no matter how much evidence accumulated—so long as we began by assigning a prior probability of zero to these hypotheses. Yet such reconsiderations provide paradigmatic examples of a rational response to the evidence. Of course, nothing in the Bayesian view *prevents* assignment of a nonzero prior probability to an hypothesis that we previously considered to be absurd; but any such assignment is a decision made in the context of discovery and all such decisions are supposed to be beyond the pale of rational analysis. In contrast, the situations described illustrate cases in which a rational response to the evidence requires that we *suspend* the mechanical application of Bayes's algorithm and return to the context of discovery. Note, however, that this return to the context of discovery requires another decision that cannot be rational on a Bayesian account.

The examples noted above provide an important perspective on the convergence theorems. At most these theorems guarantee convergence to the best supported of the hypotheses *actually under consideration*. The theorems should not lead us to expect convergence on the *correct* hypothesis unless we have somehow included that hypothesis in the set under evaluation.[1] This is another point that holds for any confirmation theory, but it is particularly poignant for Bayesians since they deny any rational basis for expanding the set of hypotheses under consideration. Investigators who begin with different sets of hypotheses (which may include some overlaps) may well converge on different final hypotheses on the basis of the same evidence. It seems reasonable to suggest that in such cases each investigator should expand the set of hypotheses under evaluation to include those that are winning the competition for a high posterior probability in the hands of their competitors. Such a decision might even lead to a more reasonable application of Bayes's law than would an insistence on continuing to gather evidence and evaluate the currently favored set of hypotheses. However, nothing in the mainstream Bayesian view requires this expansion of the set of hypotheses under consideration.[2] Indeed, if *rational evaluation* of scientific hypotheses consists *solely* in accumulating evidence and applying Bayes's law, it will be neither rational to expand the set of hypotheses nor irrational to refuse to expand this set.

Before considering further examples, let us note one possible Bayesian response to the cases mentioned thus far: that we do not really accord a zero probability to any hypothesis since there will always be some odds at which we would bet on that hypothesis. This is a dubious response for two reasons. First, there is simply no reason why we should not consider some hypotheses impossible; after all, estimates of prior probabilities are

[1]See Suppes (1966, 29–30) for a related point. Suppes argues that the Bayesian view provides no mechanism for introducing a new concept into our analysis of a body of data, and he constructs an example that leaves a Bayesian unable to choose the correct hypothesis from a set of three competing hypotheses despite how much evidence is accumulated.

[2]The claim that we ought to add any hypothesis that is seriously under consideration in the scientific community to our own set of competing hypotheses forms the centerpiece of Shimony's (1970) "tempered personalism". Shimony presents this claim as part of an account of the notion of "probability", but it is more useful to view it as a methodological proposal. Note that while this may well be a sensible proposal, any attempt to defend it as rational will require resources beyond those provided by Bayes's law and the data. Yet whether we accept Shimony's proposal can affect how we apply Bayes's law, and thereby affect the outcome of a Bayesian evaluation of specific scientific hypotheses. I return to the issue of how we choose methodological rules in sec. 4.

subjective estimates that require no justification. Second, we must not forget that any prior probability accorded to hypotheses that are considered wild must be at the expense of some probability that could otherwise be accorded to hypotheses considered more reasonable. We could, of course, distinguish between reasonable and unreasonable prior probability estimates (see Lipton 1992), but to do so is to admit that application of Bayes's law is, at most, only a part of the process of rational evaluation.

Let us consider another situation in which evidence will have no import on a Bayesian account, but in which it seems eminently rational that we redistribute prior probabilities. Suppose we are comparing two deterministic hypotheses, h_1 and h_2, to which we have assigned very different prior probabilities, $P(h_1)$ and $P(h_2)$. Now suppose we deduce an evidence statement e from one of these hypotheses, verify e, and then deduce e from the competing hypothesis. It follows from Bayes's law that the relative degree of confirmation of the two hypotheses will equal the ratio of their prior probabilities: $P(e|h) = 1$ for each hypothesis and $P(e)$ must be the same for both. As a result,

$$P(h_1|e)/P(h_2|e) = P(h_1)/P(h_2),$$

and the ratio of the probabilities is unchanged by the evidence. Bayesians maintain that this result is appropriate since nothing in this case will distinguish the two hypotheses. However, suppose we find an expanding set of cases in which we succeed in deducing a result from one of our hypotheses, verify that result, and then deduce the same result from the competing hypothesis. We may reasonably maintain that the ratio of the two posterior probabilities should approach one; yet, this cannot happen if our evaluation of the import of evidence is limited to applying the Bayesian formula. Again, nothing in the Bayesian viewpoint prevents us from suspending application of the formula and reconsidering our assignment of prior probabilities; but neither the decision to suspend rote application of Bayes's law nor the redistribution of priors will count as a rational response to the evidence on the Bayesian view.

Consider one more example which requires another form of Bayes's law that is equivalent to (1), but that makes the role of competing hypotheses explicit:

$$P(h_j|e) = P(h_j)P(e|h_j)/\Sigma P(h_i)P(e|h_i), \quad (2)$$

where the summation is taken over all of the hypotheses under consideration. Earman, with an uncharacteristic lapse of caution, says that when Bayes's law is written in this form it "shows how the probability of a hypothesis is boosted by evidence that eliminates rival hypotheses" (1992, 79; but see p. 164 where Earman expresses a different view). The idea is that if one of the hypotheses is eliminated, the denominator is reduced and the probability of each of the remaining hypotheses is raised. However, this is not the only possible outcome. We could, for example, choose to enter the context of discovery and redistribute priors. Indeed, here is one case in which this option appears mandatory: If, after eliminating the failed hypothesis, we are left with a set of hypotheses such that each entails e, and we have included a catchall hypothesis in the denominator, then the denominator must be one (see the discussion in sec. 3). In this case we had better redistribute the excess prior probability that has just become available. However, this takes us back into the context of discovery where there is no guarantee that any given prior will be raised; as a matter of principle, we are starting over. Some prior probabilities may be raised, others lowered, some dropped to zero, and new hypotheses may be introduced.

The examples discussed in this section support the main thesis stated in the final paragraph of section 1: They illustrate cases in which a rational response to the evidence requires that we do something other than continue the application of Bayes's law. In other words, even in cases in which Bayes's law could be applied, applications of this law are not automatically rational. Indeed, the examples we have been considering justify a *stronger thesis*. It is not just that applications of Bayes's law cannot be the whole of scientific rationality; applications of this law cannot even be the most fundamental part of an account of reason because, as the examples show, being rational requires that we use Bayes's law in a judicious manner. A rational agent will not mechanically apply the law to a predetermined set of hypotheses. Sometimes a rational agent will decide to suspend application of Bayes's law and return to the context of discovery in order to reconsider the set of hypotheses to be evaluated and to redistribute prior probabilities.

We can get further insight into this need to use Bayes's law in a judicious manner when applying it to confirmation theory by comparing a different application of the law. A common statistical application of Bayes's law occurs when we have a set of populations of known composition, which can be conveniently modeled by urns. We have a sample taken from one of these urns and we wish to calculate the probability that the sample came from each urn. Each of our hypotheses states that the sample came from a specific urn and we can often be confident that we know which hypotheses are relevant exactly because we are considering a definite set of populations. However, when we use Bayes's law in a confirmation theory, we have only one urn—nature, or some aspect of nature—and we are comparing hypotheses about the composition of that urn. In this case it is much more difficult to be reasonably sure that we have formulated a complete set of hypotheses. Contemporary studies of scientific revolutions have provided us with myriad examples of situations in which surprising new hypotheses came to be entertained, including cases in which new hypotheses were introduced in response to new evidence. Yet, on a Bayesian view, all decisions to include a new hypothesis in the set under evaluation amount to nonrational decisions to alter our nonrational assignments of prior probabilities and then to begin the process of empirical testing anew. This yields an extremely troubling view of the development of science. If our choice of hypotheses and distribution of prior probabilities is itself nonrational, then we have no *reason* to consider the set of hypotheses that we have been examining to be appropriate and no *reason* not to alter our hypotheses and prior probabilities at will. Indeed, if decisions to return to the context of discovery and carry out such alterations are also nonrational, then such decisions can presumably take place at any time—again, no reasons are required. Thus, as Salmon (1988, 9–10; 1990, 200) has noted, even if we have encountered convergence to one of the hypotheses in the set under consideration, we can choose to override this result by returning to the context of discovery and changing the mix of hypotheses and prior probabilities. On Bayesian grounds this will not be a rational decision, but it will be no less rational than the original decision to evaluate a particular set of hypotheses rather than some competing set. That is, the view that rational evaluation of hypotheses is completely captured in applications of Bayes's law actually leaves us with no reason why we should accept the outcome of these applications.

I do not offer the above conclusion as an instance of the ultimate nonrationality of science, but rather as an illustration of the limits of the Bayesian account of reason. To the extent that Bayesians limit scientific rationality to the application of Bayes's law, their account of the rationality of science is seriously incomplete in at least two respects: It fails to account for cases in which evidence suggests that we reconsider our choice of hypotheses and prior probabilities, and it fails to provide a reason why we should not reconsider our set of hypotheses and prior probabilities whenever we feel like doing so.

3. The Probability of the Evidence. When Bayes's law is expressed as in equation (1), $P(e)$ appears to be of considerable importance since the posterior probability of an hypothesis will depend on the value of this term. The discussion of section 1 suggests that $P(e)$ provides a second point at which a subjective estimate of a probability is required in order to apply Bayes's law. It also seems that there are two points at which evidence is brought to bear on our hypotheses: the estimate of $P(e)$ and the evaluation of $P(e|h)$. Yet this view is strikingly at variance with that of Bayesian statisticians who adhere to "the likelihood principle":

> [In] calculating $p(\theta|x)$, our inference about θ, the only contribution of the data is through the likelihood function, $p(x|\theta)$ [where θ is playing the role of our h and x the role our e]. In particular, if we have two pieces of data x_1 and x_2 with the same likelihood function, $p(x_1|\theta) = p(x_2|\theta)$, the inferences about θ from the two data sets should be the same. (Lindley 1976, 361)

Berger (1985, 28) states the point thus, "*In making inferences or decisions about θ after x is observed, all relevant experimental information is contained in the likelihood function for the observed x*". We should, then, more closely examine the role of $P(e)$ in Bayesian confirmation theory.

The first point to note is that whenever we have an exhaustive set of mutually exclusive hypotheses then, by a theorem of the probability calculus,

$$P(e) = \Sigma P(h_i)P(e|h_i). \tag{3}$$

Thus if we believe that we have formulated all of the relevant hypotheses, assigned their prior probabilities, and calculated the probability of the evidence conditional on each hypothesis, $P(e)$ is already determined. No additional estimate is to be made and we have no second point at which evidence is brought to bear on our hypothesis. In this case the supposedly nonrational decision as to which hypotheses are to be considered, along with the equally nonrational assignments of prior probabilities to these hypotheses, determines the value of $P(e)$.

Suppose, however, that we allow for the possibility that all of our explicitly formulated hypotheses may be wrong and reserve some prior probability for this option. In this case $P(e)$ appears to play an independent role. Let us rewrite (3) in a way that makes this option explicit. Using $-h$ to symbolize the case in which all of our explicit hypotheses are wrong and letting j range over the explicitly stated hypotheses, (3) becomes

$$P(e) = \Sigma P(h_j)P(e|h_j) + P(-h)\,P(e|-h). \tag{4}$$

Here some estimate for $P(e)$ would seem to be in order. But

$$P(-h) = 1 - \Sigma P(h_j). \tag{5}$$

Thus in equation (4) $P(e)$ and $P(e|-h)$ are the only undetermined parameters. As a result, assigning a value to $P(e)$ is equivalent to assigning a value to $P(e|-h)$. This is odd since $-h$ is not a specific hypothesis from which a value for e could be calculated; rather, $-h$ is an enormous disjunction of all possible hypotheses that are not included in our explicit set. Salmon, referring to $-h$ as "the catchall", has emphasized the extreme difficulty of assigning a definite value to $P(e|-h)$, "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" (1991, 329). Earman describes the likelihood of the catchall as "literally anybody's guess" (1992, 168). However, it is apparently no more difficult to estimate this likelihood than it is to assign a value to $P(e)$. Salmon (1990, 188) has noted that $P(e)$ is related to the other probabilities under consideration but has not explored the consequences of this relation. Let us explore this situation further, first for deterministic hypotheses and then for statistical hypotheses.

In the case of deterministic hypotheses we have a thoroughly bewildering situation. Suppose our explicit set of hypotheses consists of Newton's and Einstein's theories of gravity and that e is an evidence statement that is entailed by both. Then $-h$ consists of the disjunction of all hypotheses other than these two. It would be surprising if a particular e could be deduced from this disjunction since this would entail that e can be deduced from each of the disjuncts. Moreover, if every admissible hypothesis entails e then each of the $P(e|h_i)$ terms in (3) will equal one and $P(e)$ must equal one. In this case the posterior probabilities equal the prior probabilities so that e is empirically useless on a Bayesian view. An analogous point holds if every hypothesis materially implies e.

One might want to limit the set to all "relevant" hypotheses but we should consider at what point hypotheses using a four-dimensional non-Euclidean spacetime became "relevant". Even the elimination of hypotheses that entail $-e$ can be problematic. Salmon (1991, 328–331) notes the problem for early Copernicans generated by the failure to observe stellar parallax but does not conclude that the theory should have been rejected. Rather, he invokes the possibility of finding auxiliary hypotheses that, in conjunction with the Copernican view, entail no observed parallax. The likelihood of the augmented theory will be one; the effect of introducing the auxiliaries will be to require additional prior probabilities in Bayes's law, that is, the prior probabilities of the auxiliaries. I, however, leave this option aside and assume that hypotheses which entail $-e$ have indeed been rejected.

Suppose, then, that some of the hypotheses included in the catchall do not entail e and that we have eliminated those hypotheses that entail $-e$. Given that we are currently considering only deterministic hypotheses, none of these hypotheses can *entail* a fractional value of the likelihood. This leaves us with two further subcases. In one case, some of the hypotheses in our disjunction have no determinate likelihood; but then how to interpret a determinate value for $P(e|-h)$ remains unclear. Alternatively, we can allow that some of the likelihoods are fractional even though the fraction is not entailed by the hypothesis. If this is the case, however, then the values of these likelihoods will not be objectively determined given e and the relevant h; they will indeed become another set of probability values that must be estimated subjectively. Of course we cannot actually estimate these likelihoods since many of the disjuncts in $-h$ will never have been explicitly formulated. Still, significant constraints will have been placed on these likelihoods once we have estimated $P(e)$. I do not know how to interpret this situation and, for the remainder of this section, I will confine discussion to statistical hypotheses.

In the statistical case the value of $P(e)$ is not uniquely determined by our choice of hypotheses and prior probabilities, but $P(e)$ is still tightly constrained by these choices. Given that $P(e|-h)$ must be somewhere in the range from zero to one, and setting $S = \Sigma P(h_i)P(e|h_i)$, it follows from (4) that

$$S \leq P(e) \leq S + P(-h). \quad (6)$$

Thus the possible values of $P(e)$ are jointly determined by the presumably objective likelihoods and subjective prior probabilities. Indeed, $P(-h)$, which gives the allowable range for $P(e)$, is completely determined by the prior probabilities in accordance with equation (5). Moreover, $P(-h)$ will presumably be small. To assign a large value to $P(-h)$ is to judge that we are not confident that our explicit set of hypotheses includes the correct hypothesis. While this judgement may be altered by applying Bayes's law to the evidence, we should not, under the circumstances, limit our research to gathering evidence and applying Bayes's algorithm; we should also be seeking more plausible hypotheses. That is, our decision to exit the context of discovery was premature.

Salmon (1990, 191–192; 1991, 329) has offered a way of avoiding the problems generated by the catchall and the $P(e)$ term. He notes that if we write equation (1) or (2) for two hypotheses, h_1 and h_2, and take the ratios of their posterior probabilities, $P(e)$ drops out; as a result, $-h$ drops out as well. Note, however, that this proposal gives up the possibility of calculating the posterior probability of specific hypotheses and will thus be unacceptable to many Bayesians (see, e.g., Earman 1992, 171–173). For those Bayesians who do wish to calculate posterior probabilities of specific hypotheses, the discussion in the present section underlines again the extreme importance of the choice of hypotheses to be compared and of their prior probabilities. To see why, note that if we are dealing with hypotheses of any genuine scientific interest, we can expect that researchers will devise empirical tests of these hypotheses that had not been thought of when the hypotheses were formulated and their prior probabilities assigned. Now Bayesians maintain that some predicted observational results may be considered more unlikely than others, that a low estimate of $P(e)$

indicates that we consider the outcome described by e to be unlikely, and that the observation of a predicted result that we consider unlikely should provide more support for an hypothesis than an observation that is less surprising. Equation (1) shows that this happens when we use Bayes's law (ibid., 195–198; see also Niiniluoto 1983, 377). Yet we have just seen that in the case of statistical hypotheses, the possible values that we can assign to $P(e)$ are determined by e in conjunction with the set of hypotheses under consideration and the prior probabilities we have assigned to these hypotheses. Moreover, the permissible range of variation in $P(e)$ is equal to $P(-h)$, and is thus determined by our choice of prior probabilities, irrespective of any specific e that may, someday, come to be considered. Indeed, if we think that we already have an exhaustive set of hypotheses, then no empirical result can be any more surprising than any other. This result may make sense to anyone who accepts logical omniscience with a vengeance, that is, who assumes that having formulated an hypothesis we automatically know all conclusions that follow in conjunction with our background knowledge. Thus the only surprising outcomes will be those that cannot be deduced from our explicit set of hypotheses. But a methodology that assumes this degree of logical omniscience will tell us nothing about how we ought to evaluate hypotheses in real science.

If the result of this discussion seems odd, then Bayesians might want to reconsider the claim that $P(e)$ is a measure of the degree to which e is to be expected. However, what role $P(e)$ plays in Bayesian confirmation theory is now unclear. Perhaps all Bayesians should accept Salmon's conclusion that the most we should expect from Bayes's law is a ranking of relative probabilities of hypotheses.

4. Iteration. Let us consider one more way in which a reasonable Bayesian evaluation of an hypothesis depends on judgements not captured in Bayes's law. This section is aimed at those Bayesians who still want to assign values to $P(e)$ and who also maintain that even when e is known to be true, we should not substitute one for $P(e)$ in Bayes's law. Now, suppose we have acquired a body of evidence e and used it to adjust the probabilities of a set of hypotheses. In effect, we now have a new set of prior probabilities to be used in further applications of Bayes's law. Is it legitimate to use *the same evidence* a second time in order to further adjust our posterior probabilities? I am not concerned here with another instance of the same observational result, but with a reuse of a single observation.

Two arguments suggest that such reuse is legitimate. First, there is *no formal difference* between using the evidence a second time with the newly calculated posterior probabilities serving as the prior probabilities, and using that evidence for the first time with those prior probabilities. Indeed, some other Bayesian may have used the evidence in question with just those priors at which we have now arrived. Presumably, on a Bayesian approach, the formalism—not who is using it—determines the rationality of the outcome.

Second, consider the special case in which we are dealing with an exhaustive set of mutually exclusive statistical hypotheses. Once we adjust the probabilities of the hypotheses then, in general, we will also change the value of the denominator in Bayes's law. Given that we now have better estimates for the probabilities of the hypotheses, we also have an improved value for $P(e)$ in accordance with equation (3). This suggests that we should recalculate the posterior probabilities of our hypotheses using these improved inputs. The result will be a series of iterations that will hopefully yield rapid convergence to a final set of $P(h_i|e)$ values for a given e.

Instead of speculating further on such possibilities, let us determine the actual results of iteration in the special case under consideration. Assume that we have an exhaustive, mutually exclusive set of statistical hypotheses, $\{h_i\}$, and that there are k members in the set. Let $P_0(h_i)$ be the initial estimate of the probability of h_i. The value of $P(e|h_i)$ is a constant for a given e and h_i; set $P(e|h_i) = A_i$. The following expression gives the value of $P_n(h_i)$—that is, the value of $P(h_i|e)$ after n iterations (the proof is in the appendix):

$$P_n(h_i) = P_0(h_i)A_i^n/[P_0(h_1)A_1^n + P_0(h_2)A_2^n + \ldots + P_0(h_k)A_k^n]. \quad (7)$$

Now let h_1 denote the hypothesis with the largest A value.[3] Divide the numerator and denominator of (7) by A_1^n, giving

$$P_n(h_1) = P_0(h_1)/\{P_0(h_1) + P_0(h_2)\,[A_2/A_1]^n + \ldots + P_0(h_k)\,[A_k/A_1]^n\}. \quad (8)$$

Each term of the form $[A_i/A_1]$ is less than one. Thus each $[A_i/A_1]^n$ will go to zero as n increases without limit, and $P_n(h_1)$ will approach one. In other words, on iteration the hypothesis with the highest likelihood will approach a posterior probability of one. By the same token, every other hypothesis in the set will approach a posterior probability of zero.

One consequence of this result is that it can yield an explicit contradiction. Suppose that two groups of researchers who are evaluating the same set of hypotheses find two items of evidence e and e^* such that e is more probable given h_1 than given h_2, but e^* is more probable on h_2

[3]If this value is shared by two or more hypotheses, the test under consideration will not distinguish among them. For present purposes they may be considered a single hypothesis and their prior probabilities can be added. This is in accord with the likelihood principle. The argument that follows can then be applied to some other item of evidence with respect to which these hypotheses yield different likelihoods.

than on h_1. In this case one group will arrive at a probability of one for h_1 and zero for h_2; the other group will arrive at a probability of one for h_2 and zero for h_1. Moreover, nothing need change if each group now becomes aware of the other group's evidence since, as we saw in section 2, further applications of Bayes's law cannot resuscitate an hypothesis that has a prior probability of zero.

Note also that this contradiction occurs only because we are dealing, in effect, with a one-urn case. If we have several urns of known composition plus a sample taken from one of those urns, and our problem is to decide the probability that the sample came from each urn, there will be no contradiction since each new sample provides a new problem and, in effect, a new set of hypotheses.

What are we to make of this result? Having recognized the degree to which one hypothesis can swamp the others on the basis of a single piece of evidence, as well as the potential for inconsistency, we could propose a variety of methodological rules to remedy the situation. For example, we can forbid the reuse of data, or we can insist that all available evidence be taken into account in any application of Bayes's law. However, to propose such rules is to acknowledge that more is involved in the rational evaluation of an hypothesis against the evidence than simply applying Bayes's law. We are led, once again, to the conclusion that rational evaluation of hypotheses against the evidence requires that Bayes's law be used in a judicious manner. Moreover, methodological proposals must themselves be evaluated and the means by which we carry out such evaluations will have to be included in a full account of scientific rationality. After all, if our methodological rules lack a rational basis, then we have no grounds for holding that we arrive at rational results through the application of those rules.

5. From Subjectivity to Judgement. If we adhere to the traditional distinction between context of discovery and context of justification, which allows only logical relations in the latter context, then the outcome of a Bayesian evaluation carried out in the context of justification is deeply dependent on decisions made in the context of discovery. Outcomes of Bayesian evaluations depend on the entire set of hypotheses that are considered relevant, the distribution of prior probabilities among these hypotheses, and perhaps on estimates of the probability that a predicted outcome will occur. This result is a direct consequence of the Bayesian attempt to limit the context of justification to deductive arguments while relegating the choice of premises, prior probabilities, and the probability of the evidence to the context of discovery.

Salmon has long advocated a different way of thinking about the evaluation of the probabilities that are to be inserted into Bayes's algorithm.

We have already noted that Salmon proposes to eliminate the $P(e)$ term by limiting the use of Bayes's law to calculations of relative probabilities between hypotheses. However, prior probabilities cannot be eliminated from a Bayesian approach; Salmon (1990, 189–191) also argues that, in at least some cases, the likelihoods cannot be calculated in a straightforward manner. In all of these cases we must make estimates, and Salmon argues that these estimates need not be subjective. Rather, these estimates amount to evaluations of the plausibility of the hypotheses being tested, estimates that are to be based on nonformal considerations such as analogies and consistency with accepted results (in the loose sense in which "consistency" is used outside of formal logic). These estimates require judgement and Salmon (1966, 1990, 1991) maintains that these judgements are more than mere subjective opinions. Indeed, Salmon points out that these judgements, not Bayes's algorithm, constitute the crux of our *rational* evaluations of scientific hypotheses. Referring to various forms in which Bayes's law may be expressed, Salmon writes, "The algorithms are trivial; what is important is the scientific judgment involved in assessing the probabilities that are fed into the equations" (1990, 201). An adequate account of rational evaluation must include a central role for nonformal judgements.

The examples discussed support Salmon's view, but they also suggest a wider scope for judgement than Salmon has considered. Salmon seeks to limit the role for judgement to the assessment of prior probabilities and likelihoods, but we have seen that Bayesian confirmation theory requires a much greater role for judgement. We must exercise judgement in deciding which hypotheses to compare. We must also exercise judgement in deciding when to apply Bayes's law to accumulating evidence and when we should suspend the mechanical application of the algorithm in order to seek new hypotheses or redistribute prior probabilities. We have even seen that the proper use of Bayes's law may require the development and evaluation of methodological rules that will govern the reuse of evidence. In general, Bayes's law can provide a useful tool in evaluating scientific hypotheses *provided the use of the law is guided by an appropriate set of judgements.*

6. Conclusion. While the main aim of this paper has been to criticize a particular version of Bayesian confirmation theory, the general point about the role of judgement in the proper use of formal methods extends well beyond this case. Indeed, the idea that nonformal judgement is an essential element in the process of rational evaluation has provided a persistent undertone in the literature for some time. Examples include Brown (1978, 1988), Elster (1983), Lugg (1985), Newton-Smith (1981), Putnam (1981), Salmon (1966), Suppes (1984), and Wartofsky (1980). The major point urged, with varying degrees of vigor, is that while the use of appropriate algorithms is an important part of the process of arriving at rational evaluations, it is only a part. Judgement is required in order to choose appropriate algorithms and to govern their intelligent application. An account of reason that omits the central role of judgement in determining the inputs to our algorithms and in determining whether and which algorithms to use will be radically incomplete. At the same time, development of an adequate account of judgement is a seriously neglected task for the theory of reason. I have attempted to develop such an account in Brown (1988). That account will be improved and the relation between judgement and the use of formalisms in human reason will be explored to much greater depth in Brown and Hooker (1994).

APPENDIX

Proof of the iteration formula (7) by strong induction on the number of iterations:
a.) Base case: $n = 1$ recaptures Bayes's law:

$$P_1(h_i) = P_0(h_i)A_i/[P_0(h_1)A_1 + P_0(h_2)A_2 + \ldots + P_0(h_k)A_k].$$

b.) Inductive case: Assume (7) is correct up to n iterations and consider one more iteration. Let

$$S = P_0(h_1)A_1^n + P_0(h_2)A_2^n + \ldots + P_0(h_k)A_k^n,$$

so that after the nth iteration for hypothesis h_i we have

$$P_n(h_i) = P_0(h_i)A_i^n/S. \quad \text{(A1)}$$

By Bayes's law, we have

$$P_{n+1}(h_i) = P_n(h_i)A_i/[P_n(h_1)A_1 + P_n(h_2)A_2 + \ldots + P_n(h_k)A_k]. \quad \text{(A2)}$$

Substitute for $P_n(h_i)$ in (A2) using the appropriate form of (A1) for each h_i. This gives

$$P_{n+1}(h_i) = [P_0(h_i)A_i^n(A_i)/S]/[P_0(h_1)\, A_1^n(A_1)/S$$
$$+ P_0(h_2)A_2^n(A_2)/S + \ldots + P_0(h_k)A_k^n(A_k)/S].$$

Clearly, S will cancel out, and a bit of algebra yields (7).

REFERENCES

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*. 2d ed. New York: Springer-Verlag.

Brown, H. I. (1978), "On Being Rational", *American Philosophical Quarterly 15*: 241–248.

———. (1988), *Rationality*. London: Routledge.

Brown, H. I. and Hooker, C. A. (1994), *Beyond Formalism: A Naturalistic Theory of Reason*. In progress.

Earman, J. (1992), *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.

Edwards, W.; Lindman, H.; and Savage, L. (1963), "Bayesian Statistical Inference for Psychological Research", *Psychological Review 70*: 193–242.

Eells, E. (1990), "Bayesian Problems of Old Evidence", in C. W. Savage, (ed.), *Minnesota Studies in the Philosophy of Science*. Vol. 14, *Scientific Theories*. Minneapolis: University of Minnesota Press, pp. 205–223.

Elster, J. (1983), *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge, England: Cambridge University Press.

Garber, D. (1983), "Old Evidence and Logical Omniscience in Bayesian Confirmation Theory", in J. Earman, (ed.), *Minnesota Studies in the Philosophy of Science*. Vol. 10, *Testing Scientific Theories*. Minneapolis: University of Minnesota Press, pp. 99–131.

Glymour, C. (1980), *Theory and Evidence*. Princeton: Princeton University Press.

Good, I. J. (1983), "46656 Varieties of Bayesians", in *Good Thinking*. Minneapolis: University of Minnesota Press, pp. 20–21.

Howson, C. (1990), "Fitting Your Theory to the Facts: Probably Not Such a Bad Thing After All", in C. W. Savage, (ed), *Minnesota Studies in the Philosophy of Science*. Vol. 14, *Scientific Theories*. Minneapolis: University of Minnesota Press, pp. 224–244.

———. (1991), "The 'Old Evidence' Problem", *British Journal for the Philosophy of Science 42*: 547–555.

Howson, C. and Urbach, P. (1989), *Scientific Reasoning: The Bayesian Approach*. La Salle, IL: Open Court.

Jeffrey, R. (1983), "Bayesian with a Human Face", in J. Earman, (ed.), *Minnesota Studies in the Philosophy of Science*. Vol. 10, *Scientific Theories*. Minneapolis: University of Minnesota Press, pp. 133–156.

Lindley, D. V. (1976), "Bayesian Statistics", in W. L. Harper and C. A. Hooker, (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science: Proceedings of an International Research Colloquium Held at the University of Western Ontario, London, Canada, 10–13 May 1973*. Vol. 2, *Foundations and Philosophy of Statistical Inference*. Dordrecht: Reidel, pp. 353–363.

Lipton, P. (1992), Review of C. Howson and P. Urbach's *Scientific Reasoning: The Bayesian Approach*. *Mind 101*: 171–175.

Lugg, A. (1985), "The Process of Discovery", *Philosophy of Science 52*: 207–220.

Newton-Smith, W. (1981), *The Rationality of Science*. London: Routledge & Kegan Paul.

Niiniluoto, I. (1983), "Novel Facts and Bayesianism", *British Journal for the Philosophy of Science 34*: 375–379.

Putnam, H. (1981), *Reason, Truth and History*. Cambridge, England: Cambridge University Press.

Salmon, W. C. (1966), "The Foundations of Scientific Inference", in R. G. Colodny, (ed.), *Mind and Cosmos: Essays in Contemporary Science and Philosophy*. Pittsburgh: University of Pittsburgh Press, pp. 135–275.

———. (1988), "Dynamic Rationality, Propensity, Probability, and Credence", in J. Fetzer, (ed.), *Probability and Causality: Essays in Honor of Wesley C. Salmon*. Dordrecht: Reidel, pp. 3–40.

———. (1990), "Rationality and Objectivity in Science *or* Tom Kuhn Meets Tom Bayes", in W. C. Savage, (ed.), *Minnesota Studies in the Philosophy of Science*. Vol. 14, *Scientific Theories*. Minneapolis: University of Minnesota Press, pp. 175–204.

———. (1991), "The Appraisal of Theories: Kuhn Meets Bayes", in A. Fine, M. Forbes, and L. Wessels, (eds.), *PSA 1990*, vol. 2. East Lansing, MI: Philosophy of Science Association, pp. 325–332.

Savage, L. (1954), *The Foundations of Statistics*. New York: Wiley & Sons.

Shimony, A. (1970), "Scientific Inference", in R. Colodny, (ed.), *The Nature and Function of Scientific Theories*. Pittsburgh: University of Pittsburgh Press, pp. 79–172.

Suppes, P. (1966), "Concept Formation and Bayesian Decisions", in J. Hintikka and P. Suppes, (eds.), *Aspects of Inductive Logic*. Amsterdam: North Holland, pp. 21–48.

———. (1984), *Probabilistic Metaphysics*. Oxford: Oxford University Press.

Wartofsky, M. (1980), "Scientific Judgment: Creativity and Discovery in Scientific Thought", in T. Nickles, (ed.), *Scientific Discovery: Case Studies*. Dordrecht: Reidel, pp. 1–20.