

DANIEL STEEL

## BAYESIAN CONFIRMATION THEORY AND THE LIKELIHOOD PRINCIPLE

**ABSTRACT.** The likelihood principle (LP) is a core issue in disagreements between Bayesian and frequentist statistical theories. Yet statements of the LP are often ambiguous, while arguments for why a Bayesian must accept it rely upon unexamined implicit premises. I distinguish two propositions associated with the LP, which I label LP1 and LP2. I maintain that there is a compelling Bayesian argument for LP1, based upon strict conditionalization, standard Bayesian decision theory, and a proposition I call the *practical relevance principle*. In contrast, I argue that there is no similarly compelling argument for or against LP2. I suggest that these conclusions lead to a restrictedly pluralistic view of Bayesian confirmation measures.

### 1. INTRODUCTION

Despite a shared commitment to using Bayes' theorem as the basis for inductive inference, Bayesian statistics and confirmation theory generally address very different questions. Bayesian statisticians are primarily concerned to use Bayes' theorem to resolve statistical inference problems, which involves finding well-motivated procedures by which to compute posterior probability distributions over sets of alternative hypotheses in various circumstances. In contrast, Bayesian confirmation theory aims to evaluate and explicate such maxims as passing a test counts in favor of a hypothesis only if the test is severe, varied evidence confirms more strongly than narrow evidence, and so on. Yet Bayesian statisticians are not silent on the topic of rules concerning relative strength of confirmation. In particular, they generally regard the *likelihood principle* (LP) as a proposition to which Bayesianism is committed. The LP has been stated in many ways, but a common formulation goes like this: all of the information an experimental outcome provides about a parameter  $\theta$  is expressed in the likelihood function.<sup>1</sup> In this essay, I aim to disambiguate this proposition, to clarify in what sense a Bayesian is

committed to it and explore the implications of these questions for discussions of Bayesian confirmation measures.

I begin by explaining how statements of the LP like the one just given can be interpreted in more than one way. I maintain that there is a strong Bayesian argument for accepting one version of the LP (what I call LP1). This argument flows from basic Bayesian principles – particularly, strict conditionalization and the proposition that one should choose acts that maximize expected utility – together with a premise concerning the relationship between evidence and decision, which I call the *practical relevance principle*. In contrast, I maintain that no similarly compelling Bayesian reasons can be provided for or against a second interpretation of the LP (what I call LP2). Developing these arguments involves a discussion of Bayesian confirmation measures, which have been a topic of some debate in the recent Bayesian confirmation theory literature. I argue that LP1 constitutes a significant restriction on the class of acceptable Bayesian confirmation measures. The case of LP2, in contrast, suggests that there is no uniquely best Bayesian confirmation measure among those that are acceptable. Rather, among the acceptable measures, distinct measures may be suited for distinct circumstances, a viewpoint that I dub *restricted pluralism*.

## 2. THE LIKELIHOOD PRINCIPLES

What does it mean to say that all of the information that data provides about a parameter is contained in the likelihood function? To begin with, it is necessary to say what a likelihood function is. Let  $\mathbf{H}$  be a set of mutually exclusive, collectively exhaustive hypotheses, and let  $H$  be a variable ranging over members of  $\mathbf{H}$ . For instance, the hypotheses in  $\mathbf{H}$  might concern the value of the  $\theta$ , which indicates the years before present that a particular archeological site was abandoned, while  $E$  is a result of a radiocarbon dating procedure. The *likelihood function*  $L(H, E)$ , then, is by definition equal to  $kP(E|H)$ , where  $k$  is any positive constant.<sup>2</sup> One natural reading of the LP is that there is no difference in evidence if there is no difference in likelihood functions. Given the definition of likelihood function, it is easy to see that likelihood functions are the same when they are proportional (cf. Birnbaum 1962, 271). Let  $E$  and  $E^*$  be two sets of data. Then the likelihood functions  $L(H, E)$

and  $L(H, E^*)$  are *proportional* exactly if there is a constant  $k > 0$  such that  $P(E|H) = kP(E^*|H)$  for all  $H$ .

Let  $c(H, E)$  indicate the evidential support or confirmation that  $E$  provides for  $H$ . The relevant sense of confirmation here is an *incremental* one, which concerns the ‘boost’ that a hypothesis would receive if the data were learned. Incremental confirmation is usually contrasted with absolute confirmation, wherein the evidence establishes the hypothesis at or beyond a specified threshold (cf. Earman 1992, 64–67). For a Bayesian, incremental confirmation means that the data, if learned, would raise the probability of the hypothesis, while absolute confirmation means that learning the data would lift the probability to or above, say, 0.5. The exact quantitative measure of incremental confirmation is a matter of some debate, and one purpose of this essay is to explore the implications of the LP for this issue. Incremental confirmation should also be distinguished from resilience, which concerns the stability of the probability of a proposition in the face of new information (cf. Skyrms 1980). I think that there are interesting connections between incremental confirmation and resiliency, but I will have nothing to say about such matters here. The term ‘confirmation’ should be read as ‘incremental confirmation’ throughout this essay except where otherwise indicated.

Given this set up, we can state the first interpretation of the LP as follows:

LP1: If the likelihood functions  $L(H, E)$  and  $L(H, E^*)$  are proportional, then for all  $H$  in  $\mathbf{H}$ ,  $c(H, E) = c(H, E^*)$ .

Some statements of the LP are very clearly LP1 (Edwards et al. 1963, 237; Barnett 1999, 188; cf. Press 2003, 35). The LP1 is also the operative interpretation in one of the most important papers ever written about the LP, namely, Alan Birnbaum’s (1962) classic essay, “On the Foundations of Statistical Inference.”

In this paper, Birnbaum proved that the LP is equivalent to the conjunction of two propositions known as the sufficiency principle and conditionality. Conditionality states, roughly, that experimental outcomes that might have occurred but did not should have no bearing on assessments of evidence. The sufficiency principle will be discussed below. Birnbaum’s proof was important because conditionality and the sufficiency principle were far less controversial at the time than was the LP.<sup>3</sup> Here is how Birnbaum formulated the LP:

If  $E$  and  $E'$  are any two experiments with the same parameter space, represented by density functions  $f(x, \theta)$  and  $g(y, \theta)$ ; and if  $x$  and  $y$  are any respective outcomes determining the same likelihood function; then  $\text{Ev}(E, x) = \text{Ev}(E', y)$ . (1962, 271)

Birnbaum's ' $\text{Ev}(E, x)$ ' signifies the 'evidential meaning' of the outcome  $x$  of experiment  $E$  with regard to the parameters in question, in this case,  $\theta$ . Thus, Birnbaum's statement of the LP involves a notion of non-comparative confirmation, very similar to that used in the formulation of the LP1 above. Moreover, just like the LP1, Birnbaum's version of the LP addresses a situation in which one is concerned to assess the relative bearing of *two sets of experimental data* (denoted by  $x$  and  $y$ ) with regard to a *single partition of hypotheses* (where each hypothesis specifies values for  $\theta$ ). And just as in the case of the LP1, Birnbaum's version of the LP says that the two experimental outcomes have the same evidential import with regard to the set of alternative hypotheses when the likelihood functions are the same, which is to say, proportional. So, it was the LP1 that Birnbaum showed is equivalent to the conjunction of conditionality and the sufficiency principle.

Given this, one might be inclined to say that the LP *just is* LP1. However, Birnbaum immediately followed the above statement of the LP as LP1 with a much more ambiguous paraphrase.

That is, the evidential meaning of any outcome  $x$  of any experiment  $E$  is characterized fully by giving the likelihood function  $cf(x, \theta)$  (which need be described only up to an arbitrary positive constant factor) without other reference to the structure of  $E$ . (1962, 271)

Notice that this paraphrase sounds rather like the generic "the likelihood function tells you all there is to know about evidence" version of the LP. In particular, it omits the important restriction to cases in which two sets of data are being considered with regard to one set of alternative hypotheses. That leads us directly to LP2.

In addition to the different-evidence-same-hypothesis situation addressed in LP1, the LP is often thought to imply something about the import of a single set of data regarding the relative merits of a pair of alternative hypotheses. For example, consider this formulation of the LP.

Within the framework of a statistical model, all of the information which *the data* provide concerning the relative merits of *two hypotheses* is contained in the likelihood ratio of those two hypotheses on the data. (Edwards 1984, 30; italics added)<sup>4</sup>

That is, if  $E$  is a set of data and  $H$  and  $H^*$  alternative hypotheses, then the ratio  $P(E|H)/P(E|H^*)$  indicates the relative confirmation that  $E$  confers upon  $H$  with regard to  $H^*$ . Presumably, this implies that when the ratio is equal to 1, the evidential relevance of  $E$  is the same for both  $H$  and  $H^*$ . This idea can be formulated in the following manner. Let  $H$  and  $H^*$  be two hypotheses. Then:

LP2: If  $P(E|H) = P(E|H^*)$ , then  $c(H, E) = c(H^*, E)$ .

Although I regard LP1 as a better rendering of the LP, there is precedent for associating both LP1 and LP2 with the likelihood principle, and the vagueness of many formulations of the LP makes it possible to argue that LP2 is part of what it entails. And there is after all little point in debating the right way to use words. The interesting question for our purposes is whether there is some reason why a Bayesian should accept either of these two propositions. Let us turn to that question now.

### 3. CONFIRMATION MEASURES AND THE LP

It is sometimes said that the LP follows from Bayes' theorem (cf. Mayo 1996, 345; Backe 1999, S354). However, it is clear that this is not true, since the LP is a claim about relative confirmation, while Bayes' theorem is a proposition solely about probabilities. That is, Bayes' theorem says something about  $P(H|E)$  but nothing about  $c(H, E)$ . In Bayesian confirmation theory, it is standard to assume that:

$$\begin{aligned} c(H, E) &> 0 \text{ if } P(H|E) > P(H), \\ c(H, E) &< 0 \text{ if } P(H|E) < P(H), \text{ and} \\ c(H, E) &= 0 \text{ if } P(H|E) = P(H).^5 \end{aligned}$$

For convenience, I will call these the *three desiderata*. Yet Bayes' theorem does not entail the LP even when conjoined with the three desiderata, since they say nothing about relative strength of confirmation in the case when the evidence supports both hypotheses. In contrast, both the LP1 and LP2 specify conditions under which evidence is claimed to confirm to the same degree. Let us consider, then, what reasons there might be for a Bayesian to accept either of these principles.

The Bayesian argument for LP1 rests on a consequence of Bayes' theorem: if the likelihood functions  $L(H, E)$  and  $L(H, E^*)$  are proportional, then  $P(H|E) = P(H|E^*)$ , for every  $H$ . The proof of this proposition is quite simple. Let  $H_a$  be an arbitrarily chosen member of  $\mathbf{H}$ , and suppose that for all  $H_i$  in  $\mathbf{H}$ ,  $P(E|H_i) = kP(E^*|H_i)$ . Then from Bayes' theorem, we have:<sup>6</sup>

$$\begin{aligned} P(H_a|E) &= \frac{P(H_a)P(E|H_a)}{\sum_i P(H_i)P(E|H_i)} = \frac{P(H_a)kP(E^*|H_a)}{\sum_i P(H_i)kP(E^*|H_i)} \\ &= \frac{P(H_a)kP(E^*|H_a)}{k\sum_i P(H_i)P(E^*|H_i)} = \frac{P(H_a)P(E^*|H_a)}{\sum_i P(H_i)P(E^*|H_i)} \\ &= P(H_a|E^*) \end{aligned}$$

Thus, when the likelihood functions  $L(H, E)$  and  $L(H, E^*)$  are proportional, there is no difference in the posterior probability distributions. From this, LP1 follows provided we assume that:

$$(C) \quad \text{If } P(H|E) = P(H|E^*), \text{ then } c(H, E) = c(H, E^*).$$

This premise seems so obvious from a Bayesian perspective that it is likely to be assumed without mention. However, (C) follows from the three desiderata only in the special case in which  $P(H|E) = P(H)$ . In a subsequent section, I examine what reasons there are for a Bayesian to accept (C).

Let us consider, then, the Bayesian case for LP2. Suppose we are concerned with the relative bearing of the data  $E$  on two hypotheses  $H$  and  $H^*$ , which we can represent in our formalism by the ratio  $c(H, E)/c(H^*, E)$ . The reasoning begins with the premise that only those factors that can influence the ratio  $P(H|E)/P(H^*|E)$  should make a difference to  $c(H, E)/c(H^*, E)$ . Now from Bayes' theorem, we have:

$$\frac{P(H|E)}{P(H^*|E)} = \frac{P(H)P(E|H)}{P(H^*)P(E|H^*)}.$$

From this equation it can be easily seen that when  $P(E|H) = P(E|H^*)$ ,  $P(H|E)$  can differ from  $P(H^*|E)$  only if  $P(H)$  is not equal to  $P(H^*)$ . But any difference between  $P(H)$  and  $P(H^*)$  would presumably be due to some prior information, and not to  $E$ . Hence, it seems that differences between the prior probabilities of  $H$  and  $H^*$  should make no difference to the question of whether *just this evidence*  $E$  confirms  $H$  more strongly than  $H^*$ . Thus, if

the ratio of the likelihoods,  $P(E|H)/P(E|H^*)$ , equals one, then so should  $c(H, E)/c(H^*, E)$ , which is LP2.

The question of whether a Bayesian should regard either of these two arguments as sound is closely tied up with something called *confirmation measures*. A confirmation measure provides an ordering of  $c(H, E)$ , for different  $H$ 's and  $E$ 's. Given a confirmation measure, one could address such questions as whether  $E$  supports  $H$  more strongly than  $H^*$ , whether  $E$  and  $E^*$  confirm  $H$  equally, and so forth. All of the confirmation measures we will consider in fact assign specific numbers to  $c(H, E)$ , but it is only the ordinal rankings that matter. Two measures that disagree about the precise numbers, but which always agree about the orderings will be judged to be equivalent.

Let us call a measure of confirmation *Bayesian* if it satisfies the three desiderata: confirmation is positive when the data raise the probability of the hypothesis, negative when they lower the probability of the hypothesis, and neutral or irrelevant when they make no difference to the probability. There are in fact many measures that fulfill these requirements (cf. Fitelson 1999). For example, consider these three:

$$\begin{aligned}
 d(H, E) &=_{df} P(H|E) - P(H) \\
 r(H, E) &=_{df} \log \left[ \frac{P(H|E)}{P(H)} \right] = \log \left[ \frac{P(E|H)}{P(E)} \right] \\
 l(H, E) &=_{df} \log \left[ \frac{P(E|H)}{P(E|\neg H)} \right].^7
 \end{aligned}$$

Each of these Bayesian confirmation measures entails (C) and hence LP1,<sup>8</sup> but among them only  $r$  is consistent with LP2.<sup>9</sup> Moreover,  $d, r$ , and  $l$  are far from the only Bayesian confirmation measures, and among the others are several that violate (C). For instance, consider these:<sup>10</sup>

$$\begin{aligned}
 \rho(H, E) &=_{df} P(H \& E) - P(H) \times P(E) \\
 n(H, E) &=_{df} P(E|H) - P(E|\neg H) \\
 s(H, E) &=_{df} P(H|E) - P(H|\neg E).^{11}
 \end{aligned}$$

The question of whether a Bayesian is committed to the LP1 or LP2, then, can be posed in terms of confirmation measures. For LP1, the issue is whether there is some principled Bayesian reason why any confirmation measure that violates (C) is misguided.

Likewise, are there acceptable Bayesian confirmation measures that violate the LP2?

#### 4. WHY DOES A BAYESIAN NEED TO SAY ANYTHING ABOUT CONFIRMATION MEASURES?

A Bayesian might take one of several stances on confirmation measures.

- Indifference: Bayesians have no need to take any position regarding the merits of distinct confirmation measures.
- Monism: There is one true measure of confirmation, and a central task of Bayesian confirmation theory is to figure out which one it is.<sup>12</sup>
- Restricted Pluralism: There is no single confirmation measure that is best for all purposes. Nevertheless, there are restrictions that significantly limit the field of acceptable Bayesian confirmation measures.

I defend restricted pluralism. In this section, I argue (contra indifference) that Bayesians cannot merely suspend judgment on the topic of confirmation measures. In particular, since it is impossible for an observer to record *everything*, a theory of inductive inference should provide some guidance regarding which details an observer may and may not disregard in a given context. Yet propositions of this sort invariably have consequences for what confirmation measures are acceptable. Against monism, I argue that there is no single best confirmation measure among those that satisfy (C). Specifically, I maintain that there is no general, compelling Bayesian argument for or against LP2.

Discussions of Bayesian confirmation measures often do not address the challenge posed by the indifferent perspective, but rather presume that it is important to decide upon a single best confirmation measure, or at least some relatively restricted set of measures. I think that this way of proceeding is unfortunate, because it diverts attention from a question that is fundamental to the enterprise. *Why is it important to take a stance regarding the merits of distinct confirmation measures at all?* In other words, why not adopt the indifferent perspective? An adequate answer to this question involves specifying some problem that a Bayesian is obliged to address, a problem whose solution has unavoidable implications about which confirmation measures are

acceptable. But as long as this problem is left unspecified, the point of insisting on a unique measure, or even a small set (say,  $d$ ,  $r$ , and  $l$ ), is unclear. Let us consider what answer might be given to this challenge.

One possible suggestion is that Bayesian confirmation theory is committed to evaluating proposed rules of scientific method and that this cannot be done without making claims about the relative merits of confirmation measures. The Bayesian confirmation literature is filled with attempts to explicate and justify such common-sense rules of scientific method as diverse evidence confirms better than narrow evidence, severe tests are required for strong confirmation, and so forth. Moreover, it has been observed that the validity of such accounts often depends upon a particular choice of confirmation measure (cf. Fitelson 1999). So, it might be claimed that agreement on the right confirmation measure, or at least a restricted set of measures, is required for Bayesian evaluations of (purported) rules of scientific method.

This argument presupposes that Bayesian evaluations of methodological rules can go forward only if one, or a relatively small number, of confirmation measures are agreed upon. Although many assessments of rules of scientific method in the Bayesian confirmation literature do proceed by attempting to show that one type of evidence confirms better than another under certain circumstances, it is far from clear that this is the only or best strategy for a Bayesian to pursue. For instance, suppose one could show that, in certain circumstances, varied evidence is necessary for convergence of opinion among Bayesian agents. Then one might propose that the methodological value of varied data stems from its ability to create evidence driven consensus from an initial state of disagreement. Yet such an argumentative strategy is entirely independent of claims about confirmation measures. To say that data of certain types facilitate convergence, and are to be preferred for that reason, entails nothing about degrees of confirmation. A Bayesian who was utterly indifferent with regard to confirmation measures could pursue such an explanatory strategy. So, the desire to assess rules of scientific method is a good argument against the indifferent perspective only given the unsubstantiated premise that there are no other Bayesian strategies for evaluating alleged rules of scientific method. Is there some other reason why a Bayesian cannot simply be indifferent about confirmation measures?

Consider a scientist recording the results of an experiment. Such a record can be given in greater or lesser detail: more detail means more work, but an excessively abbreviated description might leave out something important. This scenario is unavoidable for any human who wishes to learn from experience. Some things are taken note of while others are disregarded, and that leads to the inevitable challenge of deciding which details matter and which do not. Any theory of inductive inference, then, ought to provide some guidance about what aspects of data can be ignored and which should be recorded. Furthermore, there is an immediate connection between confirmation and relevance: if some data  $E$  is irrelevant or does not matter to  $H$ , then  $E$  neither confirms nor disconfirms  $H$ , that is,  $c(H, E) = 0$ . Likewise, if the difference between  $E$  and  $E^*$  is irrelevant or does not matter to  $H$ , then  $E$  and  $E^*$  confirm  $H$  equally, that is,  $c(H, E) = c(H, E^*)$ . These are aspects of any plausible sense of “confirmation.” It would be bizarre to say, for instance, that  $E$  is irrelevant to  $H$  but that  $E$  nevertheless confirms  $H$ .

These simple observations concerning the relation between relevance and confirmation are illustrated by one of the central disputes among the Bayesian and Neyman–Pearson schools of statistical inference. According to the Neyman–Pearson theory, stopping rules – that is, rules specifying when the researcher will cease collecting data and commence analyzing it – are highly relevant to what inferences can be legitimately drawn from statistical data. From the Bayesian perspective, stopping rules are, aside from a few rare cases, generally irrelevant.<sup>13</sup> The Bayesian argument for the irrelevance of stopping rules is based upon LP1. For example, consider an experiment consisting of a sequence of independent and identically distributed<sup>14</sup> binary observations, such as flipping a coin, in which the outcomes are labeled by 1s or 0s. Then one stopping rule might specify that the researcher stop after 100 observations are made, while an alternative rule specifies that the researcher stop after 50 1s have been observed. Let  $E$  and  $E^*$  be the data collected in accordance with these two stopping rules, respectively, and suppose that both datasets consist of 100 observations 50 of which are 1s. Suppose that the rival hypotheses specify the probability of obtaining a 1 on a given observation. Then  $P(E|H)$  equals  $p^{50}(1-p)^{50}$  multiplied by a constant (call it  $a$ ), where  $p$  is the probability of obtaining a 1 on any observation according to  $H$ . In this example,  $a$  is the number of possible ways of getting 50 1s in 100 observations. The case of  $P(E^*|H)$  is the same except that  $p^{50}(1-p)^{50}$  is multiplied

by a different constant (call it  $b$ ). That is, given the second stopping rule, the hundredth observation had to be a 1, so  $b$  is the number of ways of getting 49 1s in 99 observations. Thus, where  $k = a/b$ ,  $P(E|H) = kP(E^*|H)$  for every  $H$ . The LP1, therefore, entails that the difference in stopping rules makes no difference to confirmation in this case. Moreover, it is easy to see that the same conclusion will ensue in any example involving a series of independent and identically distributed outcomes. In such a case, differences in stopping rules affect likelihoods only by changing the number of ways the outcome in question could have been attained, a difference that can always be expressed by a positive constant.

Notice that the general proposition concerning relevance and confirmation stated above was that no relevant difference in the data entails no difference in confirmation. However, the converse – that no difference in confirmation entails no relevant difference whatever – is not so obviously true, since the data might differ in some important way that does not affect confirmation. For example, it was noted in Section 2 that confirmation and resilience (the stability of a proposition's probability in the face of new information) are distinct. Hence, it is conceivable that a difference between two sets of data might matter for resilience but not for confirmation. But if one supposes that only evidential relevance is of concern in disputes about stopping rules, then the LP1 provides a Bayesian rationale for regarding these rules as irrelevant. The supposition that any difference in data is relevant to confirmation or else not relevant at all is also reasonable with regard to the sufficiency principle.

Consider an experiment consisting of a sequence of binary outcomes, like the ones considered above. Given this set up, the most detailed description of the experimental outcome is simply the complete sequence of the values of the variables (e.g. 10110010...). Let  $E$  represent this description of the outcome. In contrast, let  $E^*$  be some more abbreviated description of the outcome, for instance, the *number of 1s*. The sufficiency principle then asserts the following.<sup>15</sup>

- (S) If  $P(E|E^* \& H) = P(E|E^*)$ , then  $c(H, E) = c(H, E^*)$ .

When the antecedent of (S) is satisfied,  $E^*$  is said to be a *sufficient statistic*. So, the principle says that if  $E^*$  is a sufficient statistic, then recording it rather than  $E$  makes no difference to evidence. For instance, the number of 1s is a sufficient statistic when the outcomes are independent and identically distributed. The sufficiency principle,

therefore, justifies disregarding certain types of information when collecting data – in this case, the exact order of the outcomes. Like the LP1, (S) can be easily derived from Bayes' theorem, *provided that (C) is presumed*.<sup>16</sup> Moreover, the three confirmation measures listed above that violate (C) – namely,  $\rho$ ,  $n$ , and  $s$  – also violate the sufficiency principle.<sup>17</sup>

Contrary to the indifferent perspective, therefore, a Bayesian cannot simply suspend judgment about confirmation measures. For any adequate theory of inductive inference should provide guidance about which aspects of observed outcomes are relevant and which can be ignored, and Bayesian principles of this sort significantly restrict the range of acceptable confirmation measures. For instance, two of the most important rules of this kind in Bayesian statistics – LP1 and (S) – tacitly presume (C), a proposition that significantly restricts the class of admissible confirmation measures. But is there some general, compelling reason why a Bayesian must accept (C)?

## 5. THE ARGUMENT FOR (C)

According to (C), a difference in data that makes no difference to the posterior probability distribution is irrelevant, does not matter to the hypotheses in question from an evidential standpoint. But what is it for some information to *matter* for evidence? There is a very straightforward and practical way to interpret this notion. Let us say that a *decision depends on H* exactly if which action should be chosen varies according to whether  $H$  is true or false. If  $H$  is true, you should choose one thing; if it is false, something else. A given bit of information *matters* or is *relevant* to  $H$  only if it can make a difference to decisions that depend on  $H$ .

This line of reasoning suggests an immediate Bayesian argument for (C). Information that makes no difference to posterior probabilities can have no effect upon calculations of expected utility, and hence, according to Bayesian decision theory, cannot matter for evidence. Let us consider this reasoning more carefully. The argument relies on what I will call the *practical relevance principle*, or (PRP) for short.

(PRP) If learning  $E$  rather than  $E^*$  can make no difference to decisions that depend on  $H$ , then  $c(H, E) = c(H, E^*)$ .

This proposition is a formulation of the idea that information is evidentially relevant to  $H$  only when it can influence decisions that depend on  $H$ . Following Patrick Maher (1996, 159), I interpret “learning evidence  $E$ ” to entail learning it with *practical certainty*. That is, although  $E$ ’s probability may be strictly less than 1, it is close enough to make no practical difference under the circumstances. Now consider a decision that consists of choosing among a set of alternative actions  $\{a_1, \dots, a_n\}$ . According to standard Bayesian decision theory, one should choose the action that maximizes expected utility, which is defined as follows.

$$\text{(Exp)} \quad EU(a_i) = P(H)U(a_i(H)) + P(\neg H)U(a_i(\neg H))$$

In (Exp),  $U(a_i(H))$  is the utility of performing the  $i$ th action when  $H$  is true. Given this set up, if the decision depends on  $H$ , then there is at least one pair,  $a_i$  and  $a_j$ , such that  $U(a_i(H)) > U(a_j(H))$  and  $U(a_i(\neg H)) < U(a_j(\neg H))$ . The argument, then, proceeds as follows.

For a Bayesian, today’s prior probabilities are yesterday’s posteriors. Hence,  $P(H)$ , and thereby  $P(\neg H)$ , in (Exp) derive from the earlier probability of  $H$  conditional on what was learned. In this context, then, it is convenient to write the antecedent of (C) as  $P_{\text{old}}(H|E) = P_{\text{old}}(H|E^*)$ . Since we are concerned with cases in which the evidence is learned with practical certainty, we can apply the rule of *strict conditionalization*. So, if  $E$  was learned,  $P(H) = P_{\text{old}}(H|E)$ , and if  $E^*$  was learned,  $P(H) = P_{\text{old}}(H|E^*)$ . Obviously, if  $P_{\text{old}}(H|E) = P_{\text{old}}(H|E^*)$ ,  $P(H)$  is the same in either case. Hence, in this case learning  $E^*$  rather than  $E$  can make no difference to decisions that depend on  $H$ , and so by (PRP),  $c(H, E) = c(H, E^*)$ .

I view this as a compelling Bayesian argument for (C), but let us consider whether there is any Bayesian way to circumvent it. The argument does contain a few implicit premises. In particular, it is assumed that which evidence was learned has no effect on utilities and that which action is chosen is independent of the probability of the hypothesis. Although there are some circumstances in which these premises would not be appropriate, they seem entirely innocuous here, since none of the confirmation measures under consideration allows such matters to affect confirmation. Consequently, removing these premises would simply create needless complications. Since I presume that Bayesians accept the principle of strict conditionalization and the proposition that rational decision-making is a

matter of acting to maximize expected utility, criticisms would focus on (PRP).

One objection to (PRP) is that the notion of confirmation is purely cognitive and hence disconnected from practical issues about decision-making. I am skeptical that such a purely cognitive notion of evidence is tenable. Nevertheless, if one wished, one could view the decisions in question as cognitive choices about which hypothesis to accept. Maher (1993, chapters 6, 7, and 8) develops a Bayesian theory of acceptance, according to which one should accept the hypothesis that maximizes expected cognitive utility. If the utilities in (Exp) were interpreted as cognitive utilities in Maher's sense, the argument would go through as before. Thus, the thought that evidence is cognitive in a sense that practical decisions are not poses no real challenge to (PRP).

A different concern about the argument focuses on the interpretation of 'learning  $E$ ' as learning with practical certainty. This interpretation of 'learning' is required for the application of strict conditionalization. But suppose instead that 'learning  $E$ ' only requires that the probability of  $E$  be raised though not necessarily practically certain. In this case, the new probability of  $H$  would be derived from what is known as 'Jeffrey conditionalization',<sup>18</sup> according to which:

$$P_{\text{new}}(H) = P_{\text{old}}(H|E)P_{\text{new}}(E) + P_{\text{old}}(H|\neg E)P_{\text{new}}(\neg E).$$

So when the evidence remains uncertain,  $P(H|\neg E)$  matters to the new probability of  $H$ . Thus, even if  $P(H|E)$  were equal to  $P(H|E^*)$ , learning  $E$  rather than  $E^*$  could make a difference to decisions that depend on  $H$  if  $P(H|E)$  differed from  $P(H|\neg E^*)$ . As a result, information that makes no difference to the posterior probability might nevertheless matter to decisions when data is not learned with practical certainty.

However, I think that it is quite reasonable in the present context to require that a proposition qualify as evidence only if it is practically certain. Recall that 'confirmation' here is understood to mean incremental confirmation, which concerns the impact that learning some data would have on a given hypothesis. But if those data remain in serious doubt, then it is unclear why they should be regarded as evidence that has been learned. For instance, suppose  $E$  is a proposition concerning the outcome of some experiment and that  $c(H, E)$  is significantly greater than zero (that is, if learned,  $E$  would strongly confirm  $H$ ). Yet imagine that the experiment is

inconclusive. After it has been carried out, the relevant experts judge the probability of  $E$  to have been raised but still consider  $E$  to be only somewhat more probable than not. The most natural thing to say about this situation is that  $E$  is not yet evidence, though it may become such through further experiments. Consequently, it is reasonable that, when considering the impact of learning some potential evidence, one should consider what changes in belief ought to ensue from its becoming practically certain. Indeed, I think that this is how evidence is normally construed in the Bayesian confirmation literature. For example, the problem of old evidence is founded on the idea that the old evidence  $O$  is practically certain, which entails that  $P(H|O) \approx P(H)$ , where  $H$  is any hypothesis whatever.

In sum, I think that the argument presented in this section provides a strong Bayesian case for (C), and consequently for LP1. Of course, the argument does not demonstrate that LP1 is beyond question. For example, the crucial role of the premise that rational choice consists in acting to maximize expected utility shows that one who accepted some distinct account of decision-making might consistently reject LP1. Indeed, this is the situation one finds in Neyman–Pearson statistics. This theory recommends decision rules concerning the rejection and acceptance of hypotheses in which such matters as stopping rules and censored data – irrelevant by the lights of the LP1 – matter to which choice should be made.

## 6. BAYESIANISM AND LP2

In the foregoing section, I endeavored to show that there is an argument for (C) based upon core Bayesian principles, and hence that Bayesians should indeed accept LP1. In this section, I argue that no similar Bayesian case can be made for or against LP2. The most common objection to LP2 is the so-called ‘tacking paradox,’ which points out that, according to LP2,  $E$  confirms  $H$  equally well as it confirms the conjunction of  $H$  and any irrelevant addendum.<sup>19</sup> I show that the tacking paradox cannot arise in the standard type of case treated in Bayesian statistics. Hence, even if the tacking paradox is a genuine problem, it cannot be a reason for rejecting LP2 in these typical statistical examples. As a consequence, it is doubtful that any compelling, across the board Bayesian argument against LP2 exists. On the other hand, I maintain that the argument for LP2 described in Section 3 is highly problematic. Thus, I conclude

that the prospects for a compelling, general Bayesian argument for or against LP2 are dim.

Consider an archeologist who wishes to estimate the date at which a site was abandoned. Let  $\theta$  be a random variable representing that date in years before present.<sup>20</sup> Then each hypothesis consists of an assignment of a particular value to  $\theta$ , for instance, 3000, 3001, 3002, and so on. In this case, the set of alternative hypotheses are mutually exclusive, collectively exhaustive, as well as what I call *structurally identical*. I say that two hypotheses are *structurally identical* just in case each consists of an assignment of values to the same set of random variables. Hence,  $\theta = 3000$  and  $\theta = 3001$  are structurally identical, since each assigns a value to  $\theta$ . In contrast, hypotheses concerning the values of distinct variables are not structurally identical. For instance, if  $H$  is  $\theta = 3000$  while  $H^*$  is the claim that the height of the Eiffel Tower is 1000 feet, then  $H$  and  $H^*$  are not structurally identical since they do not consist of assignments of values to the same set of random variables. Similarly,  $H$  and  $H^*$  are not structurally identical if  $H$  is  $\theta = 3000$ , while  $H^*$  is the conjunction of  $\theta = 3001$  and the claim that the height of the Eiffel Tower is 1000 feet. That is, the sets  $\{\theta\}$  and  $\{\theta, \text{Height of Eiffel Tower}\}$  are distinct. In the C<sup>14</sup> dating example, then, the problem is to assess the relative evidential impact of a body of data upon a set of mutually exclusive, structurally identical alternatives. Furthermore, it is not difficult to show that the tacking paradox cannot arise in such a case.

The tacking paradox can be formulated as follows. Suppose that  $E$  confirms  $H$ , while  $H^*$  is the conjunction of  $H$  and some entirely irrelevant proposition. Then  $P(E|H) = P(E|H^*)$ , but in conflict with LP2 many – though not all (cf. Milne 1996; Maher 2004) – have the intuition that  $E$  confirms  $H$  alone more strongly than the conjunction of  $H$  and the irrelevant proposition. However, since  $H^*$  is the conjunction of  $H$  and some irrelevant proposition,  $H$  and  $H^*$  must be consistent (otherwise the conjunct would not be irrelevant to  $H$ ). Thus, in the course of comparing mutually exclusive alternatives, one will never need to ask about the relative confirmation of some data upon a hypothesis and that same hypothesis conjoined with something irrelevant. The same point holds for a related objection to the confirmation measure  $r$ . Braden Fitelson (forthcoming) proposes the following criterion for confirmation measures:

- (\*) If  $E$  provides conclusive evidence for  $H_1$ , but non-conclusive evidence for  $H_2$  (where it is assumed that  $E$ ,  $H_1$ , and  $H_2$  are all contingent claims), then  $E$  favors  $H_1$  over  $H_2$ .

For example, consider a case of the tacking paradox in which  $E$  entails  $H_1$ , and in which  $H_2$  is the conjunction of  $H_1$  and some irrelevant addendum not entailed by  $E$ . Then  $E$  conclusively establishes  $H_1$  but not  $H_2$ . Yet since  $P(E|H_1) = P(E|H_2)$ , LP2 entails that  $E$  confirms  $H_1$  and  $H_2$  equally, which contradicts (\*). It is easy to see that the situation addressed by (\*) can only occur when  $H_1$  and  $H_2$  are compatible. For when two hypotheses are mutually exclusive, evidence that conclusively establishes one refutes the other.

However, mutual exclusivity does not suffice to eliminate the concern raised by the tacking paradox, since one might also tack irrelevant addendums on to one of a mutually exclusive pair of alternatives.<sup>21</sup> That is, suppose that  $H_1$  and  $H_2$  are mutually exclusive alternatives and that  $c(H_1, E) = c(H_2, E) > 0$ . So, if  $A$  is an irrelevant addendum, then  $P(E|H_2) = P(E|H_2 \& A)$ , and hence from LP2, we have that  $c(H_1, E) = c(H_2, E) = c(H_2 \& A, E)$ . But a person who regarded the tacking paradox as a genuine counterexample would claim that  $c(H_2, E) > c(H_2 \& A, E)$  and therefore that  $c(H_1, E) > c(H_2 \& A, E)$ . In this example,  $H_1$  and  $H_2 \& A$  are mutually exclusive but not structurally identical. It is easy to see that irrelevant addendums cannot pose a problem when one is solely concerned to assess the relative support that evidence confers upon structurally identical alternatives. For whenever one hypothesis under consideration is conjoined with some irrelevant addendum while others are not, the alternatives being compared are not structurally identical. The tacking paradox, therefore, cannot arise in contexts in which one is concerned to assess the evidential impact of some data upon a set of mutually exclusive, structurally identical alternatives.

Yet assessing the relevance of some evidence upon a partition of structurally identical hypotheses is stock and trade in Bayesian statistics and in statistics more generally. Nearly, any case of a practical application of Bayesian statistics in science would share this characteristic. For instance, consider the textbook *Bayesian Approach to Interpreting Archeological Data* (Buck et al. 1996). Chapter 2 of this volume contains a general discussion of how to model a statistical inference problem from a Bayesian perspective, and this discussion explicitly treats alternative hypotheses as assignments to distinct values to parameters (Buck et al. 1996, 20–21). That is, it

is presumed that all of the inference problems of interest in that book will involve the comparison of structurally identical alternatives. The  $C^{14}$  dating example alluded to above is a case that they discuss (Buck et al. 1996, chapter 9).

Consider, then, a person who is solely interested in utilizing Bayesian methods in scientific applications, such as those described in *Bayesian Approach to Interpreting Archaeological Data*, and who furthermore wished to presume LP2 in that context. All of these applications would raise questions of the form: to what extent does the evidence  $E$  support (or undermine)  $H$  vis-à-vis its *mutually exclusive and structurally identical alternatives*? Yet neither the tacking paradox nor instances of Fitelson's principle (\*) can arise in such a context. Therefore, for a person whose use of Bayesian statistics does not extend beyond such situations, these objections *can provide no reason whatever* to reject LP2. In fact, some statements of LP2 are explicitly phrased in terms of mutually exclusive, structurally identical alternatives (cf. Royall 1997, 24). In sum, since there is a large, scientifically important class of cases in which the objections raised against LP2 are not relevant, it is doubtful that there is any compelling, across the board argument that Bayesians should reject LP2.

However, there are many scientifically interesting cases that do not involve the comparison of structurally identical alternatives. This is most obvious when one is concerned to assess the relative merits of large-scale theories, say, Newtonian Mechanics and Einstein's General Theory of Relativity. The disagreement between these two theories is hardly a mere matter of what values to assign to certain parameters. Rather these theories disagree about the basic constituents of the world and about the meaning of physical concepts such as time, space, and mass. The alternative theories might also rely upon distinct auxiliary hypotheses when generating predictions.<sup>22</sup> Examples involving large-scale rival theories that are not structurally identical, including Newtonian Mechanics and General Relativity, are often discussed in the Bayesian confirmation literature (cf. Earman 1992, 173–180). In addition, there are cases that involve comparing two mutually consistent hypotheses. For instance, one might wish to know whether the evidence  $E$  supports incorporating  $A$  as part of the hypothesis  $H$ . This question can be construed as asking which  $E$  confirms better:  $H$  or the conjunction of  $H$  and  $A$ . In such circumstances, a Bayesian might reasonably think that, if  $A$  is an irrelevant addendum, then  $E$  confirms the

conjunction of  $H$  and  $A$  less strongly than it confirms  $H$  alone. Likewise, a Bayesian could reasonably agree with Fitelson's principle (\*). So, a defense of LP2 for cases involving structurally identical alternatives is not a defense of that principle generally.

Furthermore, the positive, argument for LP2 (which was described in Section 3) is highly problematic. That argument proceeds by canvassing the possible factors that could be relevant to  $c(H, E)/c(H^*, E)$ . The first premise states that only those things capable of influencing  $P(H|E)/P(H^*|E)$  are relevant to  $c(H, E)/c(H^*, E)$ . By Bayes' theorem, the only factors that affect  $P(H|E)/P(H^*|E)$  are the ratios  $P(E|H)/P(E|H^*)$  and  $P(H)/P(H^*)$ . But, the argument continues, since  $P(H)$  and  $P(H^*)$  merely reflect prior information while  $c(H, E)/c(H^*, E)$  concerns only the relative evidential impact of  $E$  upon  $H$  and  $H^*$ ,  $P(H)/P(H^*)$  should not influence  $c(H, E)/c(H^*, E)$ . Therefore, the relative confirmation of  $E$  upon  $H$  and  $H^*$  should be determined by the ratio of the likelihoods  $P(E|H)/P(E|H^*)$ , which leads directly to LP2.

The chief difficulty with this argument is the premise that the ratio of the priors,  $P(H)/P(H^*)$ , should have no influence on the relative confirmation of  $E$  upon  $H$  and  $H^*$ ,  $c(H, E)/c(H^*, E)$ . It is precisely this claim that is called into question by the tacking paradox and Fitelson's principle (\*). Let  $E$  be the evidence,  $H$  the hypothesis, and  $A$  the irrelevant addendum. Since  $A$  is irrelevant, we may presume that  $P(E|H) = P(E|H \& A)$  and  $P(H \& A) = P(H)P(A)$ . Hence, by Bayes' theorem we have:

$$\frac{P(H|E)}{P(H \& A|E)} = \frac{P(E|H)P(H)}{P(E|H \& A)P(H \& A)} = \frac{1}{P(A)}.$$

Given the mild assumption that  $P(A) < 1$ , it follows that  $P(H \& A) < P(H)$ . Moreover, this difference in priors is solely responsible for the ratio on the left-hand side of the equation being greater than 1. In the special case in which  $P(H|E) = 1$ , we have an example of Fitelson's principle (\*). So, anyone who regards the tacking paradox as a genuine problem or who accepts Fitelson's principle must reject the premise that the ratio of the priors is irrelevant to relative confirmation.

The argument for the LP2 motivated the premise that  $P(H)/P(H^*)$  is irrelevant to  $c(H, E)/c(H^*, E)$  on the grounds that prior probabilities reflect information that has no bearing on the evidential impact of  $E$  upon  $H$  and  $H^*$ . This line of reasoning is

advanced by Elliott Sober (1993, 52) and echoed by Peter Milne (1996, 22–23). Milne writes:

Let us suppose that two theories both entail evidence statement  $E$  and  $E$  is found to be true. The fact that  $E$  is the case is then powerless to discriminate between the two hypotheses: both entail it so there is nothing *in the nature of the evidence itself* that yields grounds on which to differentiate between them. (1996, 22; italics in original)

Since the evidence cannot discriminate between the two hypotheses, Milne reasons, it must confirm them both equally. However, Milne does not provide any precise definition of what it is for evidence to ‘discriminate between two hypotheses.’ There are several possible interpretations of this phrase, but none as far as I can see results a valid, non-question begging argument.

Evidence could discriminate between two hypotheses by confirming one and disconfirming the other. Clearly, the evidence cannot discriminate in this sense when it is entailed by both hypotheses. However, even if the evidence confirms both hypotheses, it need not confirm both to the same degree, which is precisely the issue at stake in the tacking paradox and Fitelson’s (\*). So, if discrimination means confirming one hypothesis and disconfirming the other, Milne’s argument has not established that  $c(H, E) = c(H^*, E)$  if  $P(E|H) = P(E|H^*)$ . One might say that evidence ‘discriminates between two hypotheses’ when it confirms them to different degrees. Given this interpretation, Milne’s argument is valid but obviously question begging, since the issue in question is whether evidence entailed by two hypotheses might confirm one more than the other. What other senses of ‘discriminate between two hypotheses’ might there be? One possibility is that evidence discriminates between two hypotheses only if it has distinct logical or probabilistic relationships with each. However, this condition is clearly satisfied in the tacking paradox, even if one supposes that the evidence is entailed by both hypotheses. For example,  $P(H|E) > P(H \& A|E)$ , while in examples of Fitelson’s principle  $E$  logically entails  $H$  but not the conjunction of  $H$  and  $A$ . Indeed, it seems quite reasonable to say that the evidence discriminates between two hypotheses when it entails one but not the other. In short, there appears to be no sense of ‘discriminate’ that will do the work required in Milne’s argument, while there is a plausible interpretation of that term that leads to a conflicting conclusion.

Objections that have been raised against LP2 are inapplicable to the sorts of examples typically dealt with in applications of Bayesian statistics. Meanwhile, the positive argument for LP2 relies upon the inadequately supported premise that  $P(H)/P(H^*)$  is irrelevant to  $c(H, E)/c(H^*, E)$ , a premise that is called into question by objections to LP2. The prospects of a compelling, across the board Bayesian argument for or against LP2, then, appear bleak indeed, and consequently there is ample basis to doubt the existence of a one true measure of confirmation.

## 7. CONCLUSION

This essay has endeavored to clarify the sense in which Bayesians are committed to the LP. I examined two propositions associated with this principle. The first, LP1, asserted that  $E$  and  $E^*$  support  $H$  equally when the likelihood functions  $L(H, E)$  and  $L(H, E^*)$  are proportional. I argued that there are good reasons for a Bayesian to accept this proposition. Furthermore, this result significantly restricts the space of acceptable Bayesian confirmation measures, and rules out one that has been recently advocated.<sup>23</sup> The second proposition associated with the LP, which I dubbed LP2, states that  $E$  supports  $H$  and  $H^*$  equally when  $P(E|H) = P(E|H^*)$ . In contrast to LP1, I maintained that there is no general, persuasive Bayesian argument for or against this principle. Thus, although the likelihood principle places some real constraints on Bayesian confirmation measures, it does not reduce the field to a single one.

## APPENDIX

### THE CONFIRMATION MEASURE $s$ VIOLATES (C)

Recall that  $s(H, E) = P(H|E) - P(H|\neg E)$ . Some elementary probability theory and algebra shows that  $P(H|E) - P(H|\neg E)$  is equivalent to  $(P(H|E) - P(H))/(1 - P(E))$ .<sup>24</sup> This formulation makes the conflict with (C) easy to see. When the likelihood functions  $L(H, E)$  and  $L(H, E^*)$  are proportional,  $P(H|E)$  equals  $P(H|E^*)$  and there is a positive constant  $k$  such that  $P(E^*) = kP(E)$ . Thus,  $s(H, E) = (P(H|E) - P(H))/(1 - P(E))$ , while  $s(H, E^*) = (P(H|E^*) - P(H))/(1 - P(E^*)) = (P(H|E) - P(H))/(1 - kP(E))$ . So, when the likelihood functions  $L(H, E)$  and  $L(H, E^*)$  are proportional,

$P(H|E) = P(H|E^*)$  but  $s(H, E)$  differs from  $s(H, E^*)$  whenever  $k$  is not equal to 1. That  $s$  violates (C) is also shown by the probability model provided below.

THE CONFIRMATION MEASURE  $s$  VIOLATES (S)

Recall that (S), the sufficiency principle, asserts that if  $P(E|E^* \& H) = P(E|E^*)$ , then  $c(H, E) = c(H, E^*)$ . The following is a probability model that demonstrates that  $s$  violates the sufficiency principle.

$E$	$E^*$	$H$	$p$
T	T	T	.2
T	T	F	.05
T	F	T	0
T	F	F	0
F	T	T	.2
F	T	F	.05
F	F	T	.2
F	F	F	.3

In this model, the antecedent of (S) is satisfied, since

$$\begin{aligned} P(E|E^* \& H) &= \frac{P(E \& E^* \& H)}{P(E^* \& H)} = \frac{.2}{.2 + .2} = .5 \\ &= \frac{.25}{.5} = \frac{P(E \& E^*)}{P(E^*)} = P(E|E^*). \end{aligned}$$

Yet  $s(H, E) = \frac{4}{.5} - \frac{2}{.5} = .4$ , while  $s(H, E^*) = \frac{2}{.25} - \frac{4}{.75} \approx .267$ . It is also easily seen that this is an example in which  $s$  violates (C), since  $P(H|E) = P(H|E^*) = .8$ .

NOTES

<sup>1</sup> For example, see Birnbaum (1962, 271), Savage (1962, 17), and Berger and Wolpert (1988, 19).

<sup>2</sup> For a Bayesian,  $P(E|H)$  is given by  $P(E \& H)/P(H)$ . Bayarri et al. (1988) point out that it is often unclear how a probability model that defines a likelihood function can be separated from the prior distribution. That poses a difficulty for statistical theories that rely on the likelihood function but seek to eschew prior probabilities, like  $P(H)$ . However, since my interest here is solely the relationship between Bayesianism and the LP, I will assume that conditional probabilities like  $P(E|H)$ , and hence the likelihood function, are defined in reference to a complete joint probability distribution.

<sup>3</sup> Frequentist statisticians would avoid the LP by rejecting conditionality (cf. Mayo 1996, chapter 10).

<sup>4</sup> See Royall (1997, 24–25) and Hacking (1965, 219) for similar formulations of the LP. Both Royall and Hacking indicate that they regard their formulations of the LP as equivalent to Birnbaum's. But as we will see, the LP1 and LP2 are importantly distinct.

<sup>5</sup> For example, see Howson and Urbach (1993, 117) and Fitelson (1999, S363).

<sup>6</sup> For convenience, I assume here that  $\mathbf{H}$  is countable. The same reasoning works for a continuous set of alternative hypotheses, since integration, like summation, has the property  $\int kf(x)dx = k \int f(x)dx$ , where  $k$  is constant.

<sup>7</sup> For example, see Fitelson (1999, 362) and Maher (1999, 55).

<sup>8</sup> See Steel (2003, 220).

<sup>9</sup> See Fitelson (1999, S368–369).

<sup>10</sup> A version of  $\rho$  was proposed by Carnap (1962, 360), while  $n$  was suggested by Nozick (1981, 252). The measure  $s$  is advanced by Christensen (1999) and Joyce (1999, 205) as a solution to one aspect of the old-evidence problem. See Eells and Fitelson (2000a) for an argument that  $s$  is not in fact helpful for this purpose.

<sup>11</sup> See Steel (2003, 219–221) for a demonstration that  $\rho$  and  $n$  violate (C). The appendix explains why the same is also true of  $s$ .

<sup>12</sup> The best example of this type of view is undoubtedly Milne (1996). Other Bayesians advocate their favored confirmation measures in somewhat less strident fashion (cf. Schlesinger 1995; Christensen 1999; Eells and Fitelson 2000b; Fitelson 2001).

<sup>13</sup> See Mayo (1996, chapter 10), Kadane et al. (1996a, 1996b), Howson and Urbach (1993, 241–243), and Berger and Wolpert (1988) for discussions of this issue.

<sup>14</sup> That is, each outcome is probabilistically independent of all other outcomes and the probability of getting any given outcome is the same for each observation.

<sup>15</sup> See Howson and Urbach (1993, 188–192). Unlike the LP, the sufficiency principle is accepted by Bayesian and Neyman–Pearson statistical school alike.

<sup>16</sup> See Steel (2003, 224).

<sup>17</sup> See Steel (2003, 225) for a demonstration that this is so for  $\rho$  and  $n$ . That this is also true of  $s$  is shown in the appendix.

<sup>18</sup> See Earman (1992, 34).

<sup>19</sup> For example, see Fitelson (1999, S368–S369). Discussions of the tacking paradox typically presume that  $H$  entails  $E$ . But this assumption is not necessary, since the complete irrelevance of the addendum  $A$  presumably suffices to ensure that  $P(E|H) = P(E|H \& A)$ . This construal of the tacking paradox makes it equivalent to the problem of irrelevant conjunctions (cf. Fitelson 2002; Hawthorne and Fitelson 2004).

<sup>20</sup> A frequentist statistician would think of  $\theta$  as an unknown fixed quantity or parameter, rather than as a random variable. For a Bayesian,  $\theta$  is a random variable because its values correspond to subsets of the probability space covered by the agent's probability function.

<sup>21</sup> I thank Susanna Rinard for bringing this point to my attention.

<sup>22</sup> Of course, one might argue that it is precisely for such reasons that it is difficult to provide a useful Bayesian analysis of such cases. Disagreements about auxiliary hypotheses create significant difficulties for assessing likelihoods. And

given that there is no mechanical procedure for generating the complete set of theories of gravitation, we do not even know what all the alternatives are.

<sup>23</sup> The victim is the measure  $s(H, E) = P(H|E) - P(H|\neg E)$ , favorably discussed by Christensen (1999) and Joyce (1999, 205). Joyce (personal communication) agrees that  $s$  is inadequate as a measure of incremental confirmation but maintains that it is nevertheless an acceptable measure of what he terms 'effective evidence' (cf. Joyce unpublished manuscript). Whereas incremental confirmation concerns the impact on  $H$  of learning  $E$ , effective evidence concerns the extent to which the agent's belief in  $H$  depends upon a current firm belief in  $E$ .

<sup>24</sup> See Christensen (1999, 450).

## REFERENCES

- Backe, A.: 1999, 'The Likelihood Principle and the Reliability of Experiments', *Philosophy of Science* **66**, S354–S361.
- Barnett, V.: 1999, *Comparative Statistical Inference*, 3rd edition. John Wiley & Sons, New York, NY.
- Bayarri, M., M. DeGroot and J. Kadane: 1988, 'What is the Likelihood Function?', in Gupta and Berger (eds.), *Statistical Decision Theory and Related Topics IV*, Volume I. Springer-Verlag, New York, NY, pp. 3–16.
- Berger, J. and R. Wolpert: 1988, *The Likelihood Principle*, 2nd edition. Institute of Mathematical Statistics, Hayward, CA.
- Birnbaum, A.: 1962, 'On the Foundations of Statistical Inference', *Journal of the American Statistical Association* **57**, 269–326.
- Buck, C.E., W.G. Cavanagh and C.D. Litton: 1996, *Bayesian Approach to Interpreting Archaeological Data*. John Wiley & Sons, New York, NY.
- Carnap, R.: 1962, *Logical Foundations of Probability*, 2nd edition. University of Chicago Press, Chicago.
- Christensen, D.: 1999, 'Measuring Confirmation', *Journal of Philosophy* **96**, 437–461.
- Earman, J.: 1992, *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. MIT Press, Cambridge, MA.
- Edwards, A.: 1984, *Likelihood*. Cambridge University Press, New York.
- Edwards, W., H. Lindman and L. Savage: 1963, 'Bayesian Statistical Inference for Psychological Research', *Psychological Review* **70**, 193–242.
- Eells, E. and B. Fitelson: 2000a, 'Measuring Confirmation and Evidence', *Journal of Philosophy* **97**, 663–672.
- Eells, E. and B. Fitelson: 2000b, 'Symmetries and Asymmetries in Evidential Support', *Philosophical Studies* **107**, 129–142.
- Fitelson, B.: 1999, 'The Plurality of Bayesian Confirmation Measures of Confirmation and the Problem of Measure Sensitivity', *Philosophy of Science* **66** (Proceedings), S362–S378.
- Fitelson, B.: 2001, 'A Bayesian Account of Independent Evidence with Applications', *Philosophy of Science* **68** (Proceedings), S123–S140.
- Fitelson, B.: 2002, 'Putting the Irrelevance back into the Problem of Irrelevant Conjunction', *Philosophy of Science* **69**, 611–622.
- Fitelson, B.: forthcoming, 'Relational and Non-Relational Conceptions of Evidential Support', *Synthese*.

- Hacking, I.: 1965, *Logic of Statistical Inference*. Cambridge University Press, Cambridge.
- Hawthorne, J. and B. Fitelson: 2004, 'Re-solving Irrelevant Conjunction with Probabilistic Independence', *Philosophy of Science* **71**, 505–514.
- Howson, C. and P. Urbach: 1993, *Scientific Reasoning: The Bayesian Approach*, 2nd edition. Open Court, La Salle, IL.
- Joyce, J.: 1999, *The Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge.
- Joyce, J.: Unpublished manuscript, 'On the plurality of probabilist measures of Evidential Relevance'.
- Kadane, J., M. Schervish, and T. Seidenfeld: 1996a, 'When Several Bayesians Agree that there will be no Reasoning to a Foregone Conclusion', *Philosophy of Science* **63**, S281–S289.
- Kadane, J., M. Schervish, and T. Seidenfeld: 1996b, 'Reasoning to a Foregone Conclusion', *Journal of the American Statistical Association* **91**, 1228–1235.
- Maher, P.: 1993, *Betting on Theories*. Cambridge University Press, Cambridge.
- Maher, P.: 1996, 'Subjective and Objective Confirmation', *Philosophy of Science* **63**, 149–174.
- Maher, P.: 1999, 'Inductive Logic and the Ravens Paradox', *Philosophy of Science* **66**, 50–70.
- Maher, P.: 2004, 'Bayesianism and Irrelevant Conjunction', *Philosophy of Science* **71**, 515–520.
- Mayo, D.: 1996, *Error and the Growth of Experimental Knowledge*. Chicago University Press, Chicago IL.
- Milne, P.: 1996, ' $\log[p(h|eb)/p(h|e)]$  is the One True Measure of Confirmation', *Philosophy of Science* **63**, 21–26.
- Nozick, R.: 1981, *Philosophical Explanations*. Harvard University Press, Cambridge, MA.
- Press, J.S.: 2003, *Subjective and Objective Bayesian Statistics*, 2nd edition. John Wiley & Sons, New York, NY.
- Royall, R.: 1997, *Statistical Evidence: A Likelihood Paradigm*. Chapman & Hall, London.
- Savage, L.: (ed.) 1962, *Foundations of Statistical Inference: A Discussion*. Methuen, London.
- Schlesinger, G.: 1995, 'Measuring Degrees of Confirmation', *Analysis* **55**, 208–211.
- Skyrms, B.: 1980, *Causal Necessity*. Yale University Press, New Haven, CT.
- Sober, E.: 1993, 'Epistemology for Empiricists', in P. French, T. Uehling, Jr., and H. Wettstein (eds.), *Midwest Studies in Philosophy, Volume XVII, Philosophy of Science*. University of Notre-Dame Press, Notre-Dame, IA, pp. 39–61.
- Steel, D.: 2003, 'A Bayesian Way to Make Stopping Rules Matter', *Erkenntnis* **58**, 213–227.

Department of Philosophy  
 Michigan State University  
 503 S. Kedzie Hall  
 East Lansing, MI 48824  
 USA  
 E-mail: steel@msu.edu