Must a Bayesian Accept the Likelihood Principle?

Daniel Steel

Department of Philosophy

503 S. Kedzie Hall

Michigan State University

East Lansing, MI 48824-1032

Email: steel@msu.edu

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 find well-motivated procedures by which to compute posterior probability distributions over sets of alternative hypotheses. 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 regard the *likelihood principle* (LP) as a proposition to which Bayesianism is squarely 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 θ is expressed in the likelihood function.[1] Surprisingly, Bayesian confirmation theory has paid little attention to the LP, and in my experience, those working in this field sometimes regard it with suspicion. That is unfortunate, since I think that Bayesian statistics and confirmation theory are mutually illuminating in this context.

In this essay, I explain that statements of the LP like the one just given can be interpreted in more than one way. I argue that there are compelling Bayesian reasons for accepting one version of the LP (what I call LP1), while no similarly compelling Bayesian argument can be provided for a second interpretation of the LP (what I call LP2). Explaining how this is so involves a discussion of Bayesian confirmation

[1] For example, see Birnbaum (1962, 271), Savage (1962, 17), and Berger and Wolpert (1988, 19).

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.

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? One natural reading of this statement is that if there is no difference in likelihood functions, then there is no difference in evidence. There is said to be "no difference" between two likelihood functions just in case they are proportional (cf. Birnbaum 1962, 271). More exactly, let **H** be a set of mutually exclusive, collectively exhaustive hypotheses, and let *H* be a variable ranging over members of **H**. For instance, the hypotheses in **H** might concern the value of the parameter θ. 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 $P(E \mid H) = kP(E^* \mid H)$ for all *H*. The best way to appreciate what this means is by reference to a pair of graphs like the following.

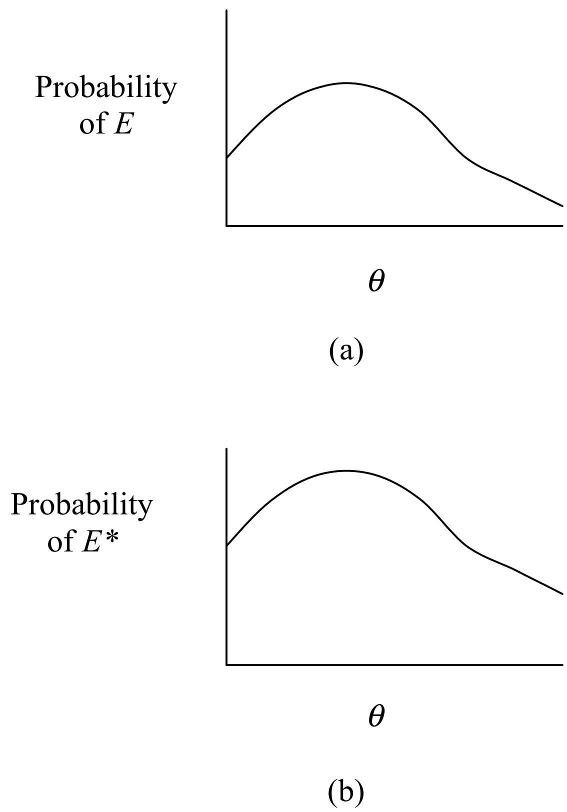

Figure 1: Proportional Likelihood Functions

Graphs (a) and (b) indicate the probability of *E* and *E**, respectively, conditional on each hypothesis in **H**. For example, the parameter θ might represent the mass of a moon of Jupiter, while *E* and *E** are distinct sets of astronomical data. Notice that the two curves are identical, except that the one in (b) is uniformly higher. So, the two likelihood functions are proportional. Let $c(H, E)$ indicate the evidential support or confirmation that *E* provides for *H*. Then 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 **H**, $c(H, E) = c(H, E^*)$.[2]

Thus, LP1 asserts that if there is no difference in the *shape* of the likelihood function, then there is no difference in evidence. Put otherwise, a difference in data whose sole effect is to move the likelihood function uniformly up or down is irrelevant as far as confirmation is concerned. Some statements of the LP are very clearly LP1 (cf. Barnett 1999, 188; Edwards et al. 1963, 237), but others suggest something rather different.

Notice that LP1 concerns cases in which one is concerned with the relative bearing of two sets of data on a given hypothesis. But in addition to this different-evidence-same-hypothesis situation, the LP is often thought 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)[3]

That is, if *E* is a set of data and *H* and *H** alternative hypotheses, then the ratio $P(E \mid H) / P(E \mid H^*)$ 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

---

[2] This rendering of the LP might be disputed on the grounds that it presumes a notion of non-comparative evidential support, *c*(*H*,*E*). In contrast, a defender of the LP might claim that evidential support is inherently contrastive, so that one cannot sensibly ask how much *E* supports *H* but only, for example, how much more *E* supports *H* than *H**. However, there is historical precedent for stating the LP in terms of a non-comparative notion of evidential support (cf. Birnbaum 1962, 271). More importantly, the substantive points about confirmation and the LP that I wish to make here are independent of this distinction. I use the "*c*(*H*, *E*)" notation because I find it convenient and because allows for easy linkages with the Bayesian confirmation literature. Those who think that confirmation is inherently contrastive can simply read "*E* and *E** confirm *H* equally" wherever they see "*c*(*H*, *E*) = *c*(*H*, *E**)."

[3] See Royall (1997, 24) for a similar formulation of the LP.

the same for both H and H^*. This idea can be formulated in the following manner. Let H and H^* be two rival hypotheses. Then:

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

I regard LP1 as a better rendering of the LP, since it captures the thought that no difference in likelihood functions means no difference in evidence. Meanwhile, LP2 does not concern likelihood *functions* but rather the likelihoods of two *particular hypotheses*. However, 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 compelling argument 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 the relationship between $P(H \mid E)$ and $P(E \mid H)$, but nothing about $c(H, E)$. In Bayesian confirmation theory, it is standard to assume that:

$$c(H, E) > 0 \text{ if } P(H|E) > P(H),$$

$$c(H, E) < 0 \text{ if } P(H|E) < P(H), \text{ and}$$

$$c(H, E) = 0 \text{ if } P(H|E) = P(H).$$[4]

For convenience, let us call these the *three desiderata*. Yet Bayes's theorem does not entail the LP even when conjoined with the three desiderata. For example, suppose that *E* and *E** both raise the probability of *H*. Then both positively confirm *H*, according to the first of the three desiderata. But these propositions say nothing about *degrees* of confirmation. Hence, they provide no basis for affirming or denying that *E* confirms *H* to the same degree as *E**. What, then, is the relationship between Bayesianism and the LP? The answer to this question evidently depends on which interpretation of the LP is at issue. Let us consider the cases of the LP1 and LP2 in turn.

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 $P(H \mid E) = P(H \mid E^*)$, for every *H*. The proof of this proposition is quite simple. Let H_a be an arbitrarily chosen member of **H**, and suppose that for all H_i in **H**, $P(E \mid H_i) = kP(E^* \mid H_i)$. Let H_a be an arbitrary member of **H**. Then from Bayes' theorem, we have:[5]

$$P(H_a \mid E) = \frac{P(H_a)P(E \mid H_a)}{\sum_i P(H_i)P(E \mid H_i)} = \frac{P(H_a)kP(E^* \mid H_a)}{\sum_i P(H_i)kP(E^* \mid H_i)} = \frac{P(H_a)kP(E^* \mid H_a)}{k\sum_i P(H_i)P(E^* \mid H_i)}$$
$$= \frac{P(H_a)P(E^* \mid H_a)}{\sum_i P(H_i)P(E^* \mid H_i)} = P(H_a \mid E^*)$$

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:

[4] For example, see Howson and Urbach (1993, 117) and Fitelson (1999, S363).

[5] For convenience, I assume here that **H** is countable. The same reasoning works for a continuous set of alternative hypotheses, since integration, like summation, has the property $\int kf(x) = k\int f(x)$, where *k* is constant.

(C) If $P(H \mid E) = P(H \mid E^*)$, 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) is not entailed by the three desiderata, as can be easily appreciated by noting that (C) provides a sufficient condition for two sets of data to positively confirm a hypothesis equally. In contrast, the three desiderata specify no such conditions. In the subsequent section, I examine what reasons there are for a Bayesian to accept (C).

Let us consider, then, the Bayesian case for LP2. The argument here amounts to pointing out a few simple features of Bayes's theorem. Suppose we are concerned with the relative bearing of the data E on two hypotheses H and H^*. Writing out Bayes' theorem, we have:

$$P(H \mid E) = \frac{P(H)P(E \mid H)}{P(E)}$$

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

From these two equations it can be easily seen that when $P(E \mid H) = P(E \mid H^*)$ the only way $P(H \mid E)$ can differ from $P(H^* \mid E)$ is 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 in 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^*.[6]

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

[6] This argument on behalf of LP2 is given by Milne (1996, 22-3).

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. 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:

$$d(H,E) =_{df} P(H \mid E) - P(H)$$

$$r(H,E) =_{df} \log\left[\frac{P(H \mid E)}{P(H)}\right] = \log\left[\frac{P(E \mid H)}{P(E)}\right]$$

$$l(H,E) =_{df} \log\left[\frac{P(E \mid H)}{P(E \mid \neg H)}\right].$$ [7]

Each of these Bayesian confirmation measures entails (C) and hence LP1,[8] but among them only *r* is consistent with LP2.[9] But *d*, *r*, and *l* are far from the only Bayesian confirmation measures, and among the others are several the violate (C). For instance, consider these:[10]

$$\rho(H,E) =_{df} P(H \,\&\, E) - P(H) \times P(E)$$

$$n(H,E) =_{df} P(E \mid H) - P(E \mid \neg H)$$

---

[7] For example, see Fitelson (1999, 362) and Maher (1999, 55).
[8] See Steel (2003, 220).
[9] See Fitelson (1999, S368-9).
[10] A version of ρ was proposed by Carnap (1962, 360), while *n* was suggested by Nozick (1981, 252). The measure *s* is advanced by Christiansen (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.

$$s(H, E) =_{df} P(H \mid E) - P(H \mid \neg E).$$[11]

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 What's a Confirmation Measure for?

There are several different stances a Bayesian might take with regard to confirmation measures.

- Indifference: Bayesians have no need for confirmation measures; all that is required is to compute posterior distributions via Bayes' theorem.
- Monism: There is one true measure of confirmation, and a central task of Bayesian confirmation theory is to figure out which one it is.[12]
- Open Pluralism: There is no single confirmation measure that is best for all purposes. Instead, there is an open ended multitude of such measures, each of which illuminates a distinct aspect of confirmation.[13]
- 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.

---

[11] See Steel (2003, 219-21) for a demonstration that ρ and n violate (C). The appendix explains why the same is also true of s.

[12] 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; Christiansen 1999; Eells and Fitelson 2000b; Fitelson 2001).

[13] See Joyce (1999, 206-7).

I will defend restricted pluralism. First, I argue (contra indifference) that there is a genuine need for assessing degrees of evidential relevance. Indeed, I claim that this need exists for any theory regarding the manner in which inductive inferences are drawn from experience, whether it is Bayesian or not. Next, contrary to open pluralism, I will argue that any Bayesian measure of confirmation that violates (C) cannot adequately meet this need. A Bayesian, therefore, is committed to LP1. Finally, against monism, I argue that there is no single best confirmation measure among those measures that satisfy (C). Specifically, I maintain that there are scientifically interesting contexts in which (D) and LP1 are reasonable, and others in which they are not.

Most discussions of Bayesian confirmation measures do not address the challenge posed by the indifferent perspective: it is assumed that a confirmation measure is needed, so the problem is just to choose which is best. I think that this way of proceeding is unfortunate, because it diverts attention from a question that is fundamental to the enterprise. *What is a confirmation measure for?* If one had a clear idea of the purpose of confirmation measures, then perhaps one could distinguish those measures capable of effectively serving that function from others. But as long as the purpose of confirmation measures is left unspecified, the point of insisting on a unique measure, or even a small set (say, *d*, *r*, and *l*), is unclear. Indeed, this is exactly the indifferent position: there is no need for a confirmation measure, so disputes about them are a waste of time. Let us consider what answer might be given to this challenge.

One possible suggestion is that the purpose of confirmation measures is to assist in the evaluation of proposed rules of scientific method. The Bayesian confirmation literature is filled with attempts to explicate and justify such commonsense 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 Bayesian confirmation measure, or at least a restricted set of measures, is required for deciding which of these accounts are sound and which not.

I do not find this to be a very convincing answer, because it presupposes that Bayesian analyses of methodological rules can go forward only if one, or a relatively small number, of confirmation measures are agreed upon. But it is far from clear that this is so. For instance, suppose one could show that, in particular types of context, varied evidence is necessary for convergence of opinion among Bayesian agents. Then one might propose that the methodological value of varied data would stem from its ability to create evidence driven consensus from an initial state of disagreement. Yet such an argumentative strategy has no need to make appeals to confirmation measures. So, the desire to assess rules of scientific method is not a compelling reason for thinking that Bayesianism has need of confirmation measures. But there are other concerns that do motivate placing restrictions on admissible 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. I suggest that these simple considerations point to a central purpose for a measure of confirmation: *to aid in distinguishing between evidentially relevant and irrelevant information*. Let E and E^* be two descriptions of the same experimental outcome, with E being the more informative of the two (so E entails E^*), and let H be the hypothesis that the experiment aims to test. If it can be shown that the same conclusions would be drawn about H from E as from E^* no matter what the experimental outcome, then there is no point exerting the extra effort involved in recording E rather than E^*. But "the same conclusions drawn about H from E^* as from E" rendered in our formalism is $c(H, E) = c(H, E^*)$.

Since (C) provides a sufficient condition for $c(H, E) = c(H, E^*)$, it specifies a sufficient condition for information being evidentially irrelevant. Not coincidentally, the LP1 is important—and controversial—precisely because of what it says does not matter for evidence. For instance, a core dispute between Bayesian and frequentist statisticians concerns stopping rules, which are a researcher's plan for when to cease collecting data and commence analyzing it.[14] According to the LP1, stopping rules are generally irrelevant, while stopping rules are very relevant from the perspective of frequentist statistics. The sufficiency principle is another commonly used statistical rule regarding what details of an experimental outcome can be safely disregarded.[15] Consider an experiment whose outcome is represented as a sequence of values of random variables. For instance, the experiment might consist in ten flips of a coin, in which case the values of the variables (Flip_1, Flip_2, etc) would heads and tails. 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. HTHHHT…). Let E represent this description of the

[14] See Mayo (1996), Kadane et al. (1996a, 1996b), Howson and Urbach (1993, 241-3), and Berger and Wolpert (1988) for discussions of this issue.
[15] See Steel (2003, 224-5).

outcome. In contrast, let *E** be some more abbreviated description of the outcome, for instance, the *number of heads*. The sufficiency principle then asserts the following.

(S) If $P(E \mid E^* \,\&\, H) = P(E \mid E^*)$, 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, granting some plausible assumptions, the number of heads is a sufficient statistic in the coin-flipping example. The important point for our purposes is that (S) can be easily derived from Bayes' theorem, *provided that* (C) *is presumed*.[16] Moreover, the three confirmation measures listed above that violate (C)—namely, ρ, n, and s—also violate the sufficiency principle.[17]

In sum, rules about which information matters and which does not are essential for empirical investigation and statistical practice. Moreover, two of the most important rules of this kind in Bayesian statistics tacitly presume (C).

5 The Argument for (C)

Let us consider, then, what Bayesian justification there might be for (C). According to (C), a difference in data that makes no difference to the posterior probability distribution does not matter, is irrelevant. But what is it for information to *matter*? 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

---

[16] See Steel (2003, 224).
[17] See Steel (2003, 225) for a demonstration that this is so for ρ and n. That this is also true of s is shown in the appendix.

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. 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 matters 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 $\{a_1, \ldots, a_n\}$. According to Bayesian decision theory, one should choose the action that maximizes expected utility, which is defined as follows.

> (Exp) $EU(a_i) = P(H)U(a_i(H)) + P(\neg H)U(a_i(\neg H))$

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 what was learned. In this context, then, it is convenient to write the antecedent of (C) as $P_{old}(H \mid E) = P_{old}(H \mid E^*)$. 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, $P(H) = P_{old}(H \mid E)$, and if E^* was learned, $P(H) = P_{old}(H \mid E^*)$. Obviously, if $P_{old}(H \mid E) = P_{old}(H \mid E^*)$, $P(H)$ 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 and evidence 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 still 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,"[18] according to which:

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

So when the evidence remains uncertain, $P(H \mid \neg E)$ matters to the new probability of *H*. Thus, even if $P(H \mid E)$ were equal to $P(H \mid E^*)$, learning *E* rather than *E** could make a difference to decisions that depend on *H* if $P(H \mid E)$ differed from $P(H \mid \neg E^*)$. 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.

But if the truth of *E* is in serious doubt, then it is unclear why it should be regarded as evidence that has been learned. Consequently, it seems reasonable that, when considering the impact of learning some potential evidence, one should consider what

[18] See Earman (1992, 34).

changes in belief ought to ensue from its becoming practically certain. In addition, one can view Jeffrey conditionalization as allowing for a generalization of the argument for (C). That is, from the above equation, it is obvious that as the new probability of *E* approaches, the old posterior probability becomes the sole determinate of the new probability of *H*.

In sum, I think that the argument presented in this section provides a very 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 what choice should be made.[19]

5 Bayesianism and LP2

In the foregoing section, I endeavored to show that (C) is criterion for any adequate Bayesian confirmation measure, and hence that Bayesians are committed to LP1. In this section, I argue that no similarly general Bayesian argument can be made for or against LP2. My strategy is to outline two scientifically interesting circumstances, one in which LP2 is reasonable and another in which it is not. I argue that LP2 is indeed a reasonable

[19] The classic argument for the LP was set out in Birnbaum (1962). Birnbaum showed that the LP followed from the sufficiency principle and a proposition called *conditionalty*, according to which experimental outcomes that might have occurred but did not should have no bearing on assessments of evidence. Frequentist statisticians would avoid the LP by rejecting conditionality (see Mayo 1996, chapter 10).

constraint for Bayesian confirmation measures when one wishes to know which hypothesis in of a set of mutually exclusive alternatives is best supported by the evidence at hand. Yet there are scientifically interesting situations in which one is concerned to evaluate a hypothesis vis-à-vis others that may be consistent with it. An example of such a case can arise when one asks which of several revisions of a hypothesis is supported by anomalous evidence. In such cases, I suggest that a confirmation measure that violates LP2 is appropriate. Thus, whereas Bayesians should accept LP1 across the board, the scope of LP2 is more narrowly restricted.

Suppose that one has a set of clearly defined mutually exclusive and collectively exhaustive alternative hypotheses. This type of situation is the standard stock and trade in Bayesian statistics and in statistics more generally. In such a context, one might have occasion to ask a question of the following sort. To what extent does *this* evidence E support (or undermine) H vis-à-vis its *alternatives*? In such a context, I maintain that the argument for LP2 described in section 3 rests upon firm ground and, consequently, that r is a reasonable Bayesian confirmation measure. I begin making this case by noting how what is often regarded as the most compelling objection to LP2 and r cannot arise when one is concerned with a set of mutually exclusive alternatives.

The most common objection to LP2 is the so-called tacking paradox. Suppose that H entails E, and suppose that H^* is the conjunction of H and some entirely irrelevant proposition. Then $P(E \mid H) = P(E \mid H^*)$, but in conflict with LP2 many have the intuition that E confirms H alone more strongly than the conjunction of H and the irrelevant proposition. However, the tacking paradox obviously requires that H and H^* are consistent (otherwise the conjunct would not be irrelevant to H). Thus, the tacking paradox can never pose a problem in situations in which one is concerned to choose among mutually exclusive alternatives. The same point holds for a related objection to *r*. Braden Fitelson (fortcoming) 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.

To illustrate (*), Fitelson asks us to consider the case in which *E* is the proposition that a spade was drawn from a deck of cards, while H_1 is the proposition that the ace of spades was drawn and H_2 that a black card was drawn. In this case, *E* favors both H_1 and H_2, but it provides conclusive evidence only for H_2. Yet according to *r*, *E* confirms H_1 more strongly than H_2, since $P(E \mid H_1)$ is greater than $P(E \mid H_2)$. Again, this is a case in which the two hypotheses being compared are logically consistent. Moreover, 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. Thus, *r* and LP2 are compatible with (*) so long as one restricts attention to situations in which mutually exclusive alternatives are being compared.

Let us turn to the positive argument for LP2 presented in section 3. That argument began with the observation that when $P(E \mid H) = P(E \mid H^*)$, $P(H \mid E)$ differs from $P(H^* \mid E)$ only if $P(H)$ is distinct from $P(H)$. Next, it was claimed that $c(H, E)$ indicates the support that *just this evidence E* confers upon *H*. This premise seems very reasonable, since the hypothesis that is best supported all things considered is, for a Bayesian, given by the posterior probability distribution. It also seems reasonable that differences in the priors reflect information and evidence other than *E*. So, there is a reason to think that when $P(E \mid H) = P(E \mid H^*)$, E supports H to the same degree as H^*. Moreover, when one is concerned to compare mutually exclusive alternative hypotheses, problem examples for LP2 are ruled out of court.[20] Hence, in this context, it seems reasonable for a Bayesian to accept the LP2, although perhaps not mandatory.

However, if one is concerned with a set of hypotheses, not all of which are mutually exclusive, then the problem examples, like the tacking paradox, might arise. The pertinent question, I think, is whether there are any scientifically interesting cases of this sort. No doubt, one can contrive examples about cards drawn from standard decks and the like. But if such cases do not correspond to any circumstance that one might seriously confront in science, it is unclear why an advocate of LP2 should be concerned with them. In fact, I think that there are scientifically interesting cases in which one is concerned with the relative support that evidence confers upon hypotheses that may not be mutually exclusive. Such cases can occur in situations that, in the philosophy of science, would be grouped under the label of the "Duhem problem."

Suppose that one is considering what modifications to make to a previously well-regarded hypothesis H in the face of some disconfirming evidence E. The Duhem problem consists in the observation that deductive logic alone is incapable of telling us just which change to make. Typically, there are several potential emendations of H to choose among, and one might even decide to not amend H at all but instead decide that the evidence is not trustworthy. Suppose that we decide that some emendation of H is called for. Then we are confronted with the following question. *Which of the many revisions of H is best supported by the anomalous evidence E*?

[20] Indeed, the assumption that the hypotheses under consideration are mutually exclusive is typically built into statements of the LP, as is the case here (see the statements of LP1 and LP2 in section 2).

One solution would be to simply toss out whatever aspect of *H* it was that led to the trouble. Call this denuded version of the original hypothesis, *H*–. Another possibility would be to not merely toss out the problematic feature, but also to replace it with something that would result in a better fit with the evidence. Such a hypothesis might simply be the conjunction of *H*– and some additional proposition *A*. In this case it is obvious that the hypotheses under consideration are not all mutually exclusive. The question, then, is which one (*H*–, *H*– & *A*, etc.) is best supported by the evidence *E*. To take an easy case, the evidence recommends *against* the addition of *A* to *H*– if $P(\mathrm{E} \mid H–) = P(E \mid H– \,\&\, A)$ and probability of the conjunction of *H*– and *A* is lower than that of *H*– alone. This is exactly the situation of irrelevant conjunctions raised by the tacking paradox. Furthermore, in this situation, it is clear that the priors *should not* be ignored. Whether the evidence supports a particular modification to the hypothesis depends in part on the antecedent plausibility of the hypothesis so modified.

The above considerations suggest that LP2 is not reasonable in cases in which one must decide which revision in a hypothesis is supported by some anomalous data. Thus, whereas LP1 is a principle that Bayesians should accept generally, LP2 is reasonable in some circumstances but not others. One might argue that the LP2 was intended to be restricted to comparisons of mutually exclusive hypotheses all along. In that case, the upshot of the above reasoning is that the confirmation measure *r*, though reasonable when comparing mutually exclusive alternatives, is inappropriate in other contexts. Given this the prospects of identifying a unique Bayesian confirmation measure that is best for all circumstances are dim.

6 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 although some confirmation measures conflict with this proposition, LP1 is something that any Bayesian should accept. This result significantly restricts the space of acceptable Bayesian confirmation measures, and rules out one that has been recently advocated.[21] The second proposition associated with the LP (which I dubbed LP2) states that E supports H and H^* equally when $P(E \mid H) = P(E \mid H^*)$. I argued that there is an important class of cases—namely, those in which one is concerned to choose among a set of mutually exclusive hypotheses—in which LP2 is quite reasonable. Nevertheless, I also argued that there are scientifically interesting contexts in which LP2 is inappropriate, for example, when one is concerned to decide which of several modifications of a hypothesis is best supported by some anomalous evidence. That LP2 is reasonable in some cases but not others shows that there is no hope of identifying a "one true measure" of Bayesian confirmation. Among those that are acceptable, distinct measures are suited to distinct circumstances.

Appendix

1. The Confirmation Measure *s* Violates (C): Recall that $s(H, E) = P(H \mid E) - P(H \mid \neg E)$. Some elementary probability theory and algebra shows that $P(H \mid E) - P(H \mid \neg E)$ is

[21] This victim is the measure $s(H, E) = P(H \mid E) - P(H \mid \neg E)$, favorably discussed by Christiansen (1999) and Joyce (1999, 205).

equivalent to $\dfrac{P(H \mid E) - P(H)}{1 - P(E)}$.[22] This formulation makes the conflict with (C) easy to see. When the likelihood functions L(*H*, *E*) and L(*H*, *E**) are proportional, $P(H \mid E)$ equals $P(H \mid E^*)$ and there is a positive constant *k* such that $P(E^*) = kP(E)$. Thus, $s(H, E)$

$$= \frac{P(H \mid E) - P(H)}{1 - P(E)}, \text{ while } s(H, E^*) = \frac{P(H \mid E^*) - P(H)}{1 - P(E^*)} = \frac{P(H \mid E) - P(H)}{1 - kP(E)}.$$

So, when the likelihood functions L(*H*, *E*) and L(*H*, *E**) are proportional, $P(H \mid E) = P(H \mid E^*)$ 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.

2. The Confirmation Measure *s* Violates the (S): Recall that (S), the sufficiency principle, asserts that if $P(E \mid E^* \,\&\, H) = P(E \mid E^*)$, 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

[22] See Christiansen (1999, 450).

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

$$P(E \mid E^* \& H) = \frac{P(E \& E^* \& H)}{P(E^* \& H)} = \frac{.2}{.2 + .2} = .5 = \frac{.25}{.5} = \frac{P(E \& E^*)}{P(E^*)} = P(E \mid E^*) .$$

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 $P(H \mid E) = P(H \mid E^*) = .8$.

References

Backe, Andrew (1999), "The Likelihood Principle and the Reliability of Experiments", *Philosophy of Science* 66: S354-S361.

Barnett, Vic (1999), *Comparative Statistical Inference*, 3rd edition. New York, NY: John Wiley & Sons.

Berger, J. and R. Wolpert (1988), *The Likelihood Principle*. 2nd edition. Hayward, CA: Institute of Mathematical Statistics.

Birnbaum, Allan (1962), "On the Foundations of Statistical Inference", *Journal of the American Statistical Association* 57: 269-326.

Carnap, Rudolf (1962), *Logical Foundations of Probability*, 2nd edition. Chicago: University of Chicago Press.

Christiansen, David (1999), "Measuring Confirmation", *Journal of Philosophy* 96: 437-461.

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

Edwards, A. (1984), *Likelihood*. New York: Cambridge University Press.

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

Eells, Ellery and Branden Fitelson (2000a), "Measuring Confirmation and Evidence", *Journal of Philosophy* 97: 663-672.

_____ (2000b), "Symmetries and Asymmetries in Evidential Support", *Philosophical Studies* 107: 129-42.

Fitelson, Branden (1999), "The Plurality of Bayesian Confirmation Measures of Confirmation and the Problem of Measure Sensitivity", *Philosophy of Science* 66 (Proceedings): S362-S378.

_____ (2001), "A Bayesian Account of Independent Evidence with Applications", Philosophy of Science 68 (Proceedings): S123-S140.

_____ (forthcoming), "Relational and Non-Relational Conceptions of Evidential Support", *Synthese*.

Howson, Colin, and Peter Urbach (1993), *Scientific Reasoning: the Bayesian Approach*, 2nd edition. La Salle, IL: Open Court.

Joyce, James (1999), *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

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.

_____ (1996b), "Reasoning to a Foregone Conclusion", *Journal of the American Statistical Association* 91: 1228-35.

Maher, Patrick (1993), *Betting on Theories*. Cambridge: Cambridge University Press.

_____ (1996), “Subjective and Objective Confirmation”, *Philosophy of Science* 63: 149-174.

_____ (1999), “Inductive Logic and the Ravens Paradox”, *Philosophy of Science* 66: 50-70.

Mayo, Deborah (1996), *Error and the Growth of Experimental Knowledge*. Chicago IL: Chicago University Press.

Milne, Peter (1996), “log[$p(h/eb)/p(h/b)$] is the One True Measure of Confirmation”, *Philosophy of Science* 63: 21-26.

Nozick, Robert (1981), *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

Royall, R. (1997), *Statistical Evidence: A Likelihood Paradigm*. London: Chapman & Hall.

Savage, L. (ed.) (1962), *Foundations of Statistical Inference: A Discussion*. London: Methuen.

Schlesinger, G. (1995), “Measuring Degrees of Confirmation”, *Analysis* 55: 208-211.

Steel, Daniel (2003), “A Bayesian Way to Make Stopping Rules Matter”, *Erkenntnis* 58: 213-227.