

DANIEL STEEL

A BAYESIAN WAY TO MAKE STOPPING RULES MATTER*

ABSTRACT. Disputes between advocates of Bayesians and more orthodox approaches to statistical inference presuppose that Bayesians must regard stopping rules, which play an important role in orthodox statistical methods, as evidentially irrelevant. In this essay, I show that this is not the case and that the stopping rule is evidentially relevant given some Bayesian confirmation measures that have been seriously proposed. However, I show that accepting a confirmation measure of this sort comes at the cost of rejecting two useful ancillary Bayesian principles.

1. INTRODUCTION

A stopping rule specifies when the researcher will cease gathering new data and commence analyzing what has been collected so far. Stopping rules play a prominent role in the disputes between advocates of the Bayesian approach to statistical inference and defenders of more orthodox methods, such as the Neyman-Pearson statistical theory. Bayesians say that stopping rules are irrelevant to statistical inference while the defenders of orthodox statistical theories take the contrary position. Deborah Mayo (1996, chapter 10) attempts to turn this difference to the advantage of Neyman-Pearson statistics by arguing that, as a result of disregarding stopping rules, Bayesians sometimes must judge it irrelevant that an experimental procedure was assured of generating an incorrect result.¹ Joseph Kadane et al. (1996a, 1996b) respond by showing that this argument works only if the Bayesian prior probability distribution violates the principle of countable additivity, which several Bayesians have argued on independent grounds is a criterion of rational degrees of belief (cf. Howson and Urbach 1993, p. 81; Maher 1993, pp. 198–9; Williamson 1999). For their part, Bayesians and other critics of orthodox methods charge that allowing stopping rules to influence statistical inference leads to intuitively absurd conclusions (cf. Pratt 1962, pp. 314–5; Kyburg 1974, pp. 53–4; Berger and Wolpert 1988, pp. 90–2; Howson and Urbach 1993; pp. 241–3).

In what follows, I demonstrate that a central presupposition of this debate is mistaken. Specifically, I show that it is possible for a Bayesian to judge stopping rules relevant to assessing the evidential significance of a



body of data, provided that an appropriate Bayesian confirmation measure is adopted. However, I also show that adopting such a confirmation measure necessitates relinquishing two ancillary principles that are useful to Bayesian confirmation theory and statistics. Thus, the reason for a Bayesian to reject the relevance of stopping rules is not that this position follows inexorably from core Bayesian principles; instead, it is that doing so would be a cause of considerable inconvenience.

2. THE LIKELIHOOD PRINCIPLE AND STOPPING RULES

Both sides of the dispute sketched in the foregoing section assume that Bayesians must dismiss stopping rules because their theory is inherently tied to the likelihood principle (LP), which entails that the stopping rule is irrelevant to the evidential force of the experimental outcome (cf. Berger and Wolpert 1988, pp. 74–89).

The likelihood of a hypothesis H given some data E is a measure of the probability of E on the assumption that H is true. In particular, suppose that \mathbf{H} is a set of mutually exclusive hypotheses such that we are certain that the true hypothesis is in \mathbf{H} . Throughout this paper, the boldfaced \mathbf{H} shall be used to represent sets of alternative hypotheses of this sort. Then for every H_i in \mathbf{H} , the likelihood of H_i given E , $L(H_i, E)$, is equal to $kP(E|H_i)$, where k is some positive constant (cf. Edwards 1984, 9–10). An immediate consequence of this definition is that $L(H_i, E) = L(H_j, E)$ just in case $P(E|H_i) = P(E|H_j)$, where H_i and H_j are any two members of \mathbf{H} . Thus, in the following discussion, I will substitute equalities of the latter sort for those of the former kind for convenience.

The LP can be stated in several different ways; I choose the formulation that is most directly relevant to our concerns. Let us introduce the notation $c(H, E)$ to represent the degree of evidential support or confirmation that E provides for H . Suppose that E and E' are two sets of data produced by two separate experiments. Then according to the LP:

LP: If there is a constant $k > 0$ such that $P(E|H_i) = kP(E'|H_i)$ for all $H_i \in \mathbf{H}$, then for all $H_i \in \mathbf{H}$, $c(H_i, E) = c(H_i, E')$.²

If we accept the above principle, then we must conclude that the stopping rule is generally irrelevant to the assessment of the evidential force of an experimental outcome.

To see how this is so, consider the following simple example. Suppose that our experiment consists in flipping the coin ten times and recording

the number of heads. The statement E might assert that five heads were obtained, and H might be the hypothesis that the coin is fair. In this case, $P(E|H)$ would be computed by the following formula:

$$C_x^n p^x (1 - p)^{n-x}$$

where x is the number of heads, n the number of flips, p the probability of heads specified by the hypothesis, and C_x^n shorthand for the quantity $n!/x!(n-x)!$. Notice that C_x^n is a positive constant, once the values of x and n are fixed. For example, in the present case we have $C_5^{10} = 252$. Let each H_i in \mathbf{H} specify some point value p_i for the probability of heads. Thus, for each H_i in \mathbf{H} , $P(E|H_i) = 252p_i^5(1 - p_i)^5$.

However, we could have designed the experiment differently so that the number of heads would be fixed in advance, say at five, while the number of flips could vary from one repetition of the experiment to the next. That is, we could plan to flip until we got five heads, and we would be concerned with how many flips were required for this result. Imagine that the fifth head turned up on the tenth flip, and let E' record this information. This is an example of a negative binomial experiment, and the likelihoods for such an experiment can be computed by the following formula:

$$C_{x-1}^{n-1} p^x (1 - p)^{n-x}$$

where x , n , and p are interpreted as in the binomial formula. Hence, for each H_i in \mathbf{H} , $P(E'|H_i) = 126p_i^5(1 - p_i)^5$. Moreover, we can easily see that $P(E|H_i) = kP(E'|H_i)$, for each H_i when $k = 252/126 = 2$. Hence, the LP tells us that, for each H_i in \mathbf{H} , the degree of evidential support given by E to H_i equals that which E' gives to H_i .

In general, then, the stopping rule affects likelihoods only in a way that can be expressed by a positive constant.³ Therefore, the LP tells us that the stopping rule is irrelevant to the assessment of the evidential significance of the experimental result. However, the difference between the binomial and negative binomial distribution does make a difference from the point of view of frequentist theories of statistical inference. For example, one can construct examples in which the difference between the binomial and negative binomial would make the difference between “accept” and “reject” in a standard significance test, even though the same data was generated in each case (cf. Berger and Wolpert 1988, pp. 20–1). Defenders of the LP take such cases as illustrations of the intuitively bizarre consequences of the frequentist approach to statistical inference.

3. THE LP AND THE BAYESIAN CORE

Let us begin with a brief exposition of the Bayesian approach to scientific inference. According to Bayesians, rational people assign degrees of belief to propositions that can be represented as probabilities. Scientific inference, then, is a matter of altering such degrees of belief as additional information is acquired. Bayes's theorem plays a central role in determining how degrees of belief are to be changed. Bayes' theorem states:

$$(1) \quad P(H_a|E) = \frac{P(H_a)P(E|H_a)}{\sum_i P(H_i)P(E|H_i)},$$

Where the H_i 's are the members of \mathbf{H} , and H_a is any H_i .

It is important to remember that the probabilities in (1) represent the degrees of a particular (ideally rational) agent at a particular time. Since Bayesian inference concerns how degrees of belief change over time, it is necessary that some further principle be given that relates an agent's beliefs at one time to his later beliefs. The most commonly used principle for this purpose is called *strict conditionalization*. Let the probability function $P_{\text{new}}(\bullet)$ represent the agent's degrees of belief after he has learned E and nothing else. Let the probability function $P_{\text{old}}(\bullet)$ represent the same agent's degrees of beliefs immediately prior to learning E . Then strict conditionalization asserts that, for any H ,

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

So far we have said nothing about confirmation or evidence. The following qualitative statements about confirmation are commonly accepted by Bayesians:

E confirms or supports H when $P(H|E) > P(H)$
 E disconfirms or undermines H when $P(H|E) < P(H)$
 E is neutral with respect to H when $P(H|E) = P(H)$.
 (Howson and Urbach 1993, p. 117)

I call these propositions "qualitative" because they provide no analysis of *degrees* of confirmation – no analysis of what it is, for example, for E to more strongly confirm H than E' . That is what a confirmation measure is supposed to provide. A confirmation measure will be said to be a *Bayesian confirmation measure* just in case it entails the three qualitative principles of confirmation stated above.

In what follows, the interpretation of probabilities as degrees of belief, the principle of strict conditionalization, and the claims about qualitative confirmation enunciated above shall be collectively referred to as the *Bayesian core*. This expression is intended to indicate that anyone who accepts these propositions can be justifiably labeled a Bayesian.⁴ What I show, therefore, is that one can consistently accept the Bayesian core while rejecting the LP.

Let us consider whether accepting the Bayesian core commits one to the LP. Mayo argues that Bayesians are saddled with the LP by citing prominent Bayesian statisticians who have endorsed the principle (1996, pp. 319, 340) and by asserting that “the LP follows from Bayes’ theorem” (ibid., p. 345).⁵ But these arguments are unpersuasive. First, that prominent Bayesians have advocated the LP does not demonstrate that Bayesian confirmation theory is inextricably linked to it. Second, it is *false* that Bayes’ theorem entails the LP. The reason for this is very simple: the LP is a principle concerning evidence, whereas Bayes’ theorem is an innocuous theorem of the probability calculus that says nothing about evidence. Therefore, Bayes’ theorem *alone* does not entail the LP.

Moreover, since the Bayesian core says nothing about degrees of confirmation, it does not entail the LP. Suppose for example that $P(H|E) > P(H)$ and $P(H|E') > P(H)$. Then the qualitative definitions of confirmation given above allow us to conclude that E and E' both confirm H . However, these definitions provide *no basis* for asserting, or denying, that E and E' each confirm H to the same degree. Thus, accepting the Bayesian core does not require taking any particular stance with regard to the LP.

However, accepting a particular confirmation measure *in addition to* the Bayesian core changes all this. Consider these three examples of commonly invoked Bayesian confirmation measures:

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

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

$$l(H, E) =_{\text{df}} \log \left[\frac{P(E|H)}{P(E|\bar{H})} \right].$$

(cf. Fitelson 1999, p. 362; Maher 1999, p. 55)

For ease of exposition, I will refer to these three confirmation measures as d , r , and l , respectively.

It can be easily shown that d , r , and l are each Bayesian confirmation measures, given the natural convention that:

$$\begin{aligned} E \text{ confirms } H & \text{ if and only if } c(H, E) > 0, \\ E \text{ disconfirms } H & \text{ if and only if } c(H, E) < 0, \\ E \text{ is neutral with respect to } H & \text{ if and only if } c(H, E) = 0. \end{aligned}$$

The case is trivial for d and r . That l is a Bayesian confirmation measure can also be seen once one notes that Bayes' theorem can be written like so:

$$(2) \quad P(H|E) = \frac{P(H)}{P(H) + \frac{P(E|\bar{H})}{P(E|H)}P(\bar{H})}.$$

However, d , r , and l are far from being the only Bayesian confirmation measures. For example, consider these three:

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

$$n(H, E) =_{\text{df}} P(E|H) - P(E|\bar{H})$$

$$m(H, E) =_{\text{df}} P(E|H) - P(E).^6$$

Again for convenience, I shall refer to these three measures as ρ , n , and m , respectively. It is easy to see that ρ is a Bayesian confirmation measure once one observes that $P(H \& E) = P(H|E)P(E)$. So ρ equals zero exactly when $P(H|E) = P(H)$; ρ is greater than zero when $P(H|E) > P(H)$, and ρ is less than zero when $P(H|E) < P(H)$. Likewise, a cursory examination of the equation in (2) shows that n is a Bayesian confirmation measure. That the same is true for m is also trivial once we note that, by Bayes' theorem,

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

The confirmation measures listed above are not equivalent, in the sense that they may produce conflicting rankings of the degree of confirmation that a body evidence confers on the members of a set of hypotheses (Fetelson 1999, S364). Not surprisingly, then, there is some dispute among Bayesians as to which measure ought to be preferred (cf. Schlesinger 1995, Milne 1996). For example, Peter Milne (1996) purports to give a demonstration that r is the "One True Measure of Confirmation". Other Bayesians

give what they take to be compelling reasons for rejecting r in favor of d or l (cf. Fitelson 1999, S368–9).⁷ The most common argument of this kind, sometimes called the “tacking paradox”, is based on what some Bayesians perceive to be an undesirable feature of r . That is, if H entails E , and H' is the conjunction of H and any arbitrary hypothesis whatsoever consistent with H , then according to r , E confirms H and H' to the same degree. Milne, for his part, does not find this consequence undesirable and asserts that “the tacking paradox is a wretched shibboleth” (1996, p. 23).

We need not be detained by these internecine squabbles. What is important for our purposes is that the different Bayesian confirmation measures often lead to divergent conclusions about confirmation (cf. Fitelson 1999). As we will see below, the LP is a case in point. In particular, I show that the confirmation measures ρ , n , and m , but *not* measures d , r , and l , are k -measures, where the expression “ k -measure” is defined as follows.

μ is a k -measure if and only if (a) μ is a Bayesian confirmation measure, and (b) if there is a constant $k > 0$ such that $P(E|H_i) = kP(E'|H_i)$ for each $H_i \in \mathbf{H}$, then for each $H_i \in \mathbf{H}$, $\mu(H_i, E) = k\mu(H_i, E')$.

Obviously, every k -measure violates the LP, so the interesting question lies in whether k -measures exist. We turn to that issue now.

4. BAYESIAN CONFIRMATION MEASURES THAT VIOLATE THE LP

The following is a theorem of the probability calculus:

LIKELIHOOD THEOREM. Let $k > 0$ be constant, and for all H_i in \mathbf{H} , let $P(E|H_i) = kP(E'|H_i)$. Then for all H_i in \mathbf{H} , $P(H_i|E) = P(H_i|E')$.

The proof of the likelihood theorem is quite simple. Let H_a be an arbitrarily chosen member of \mathbf{H} , and suppose that for all H_i in \mathbf{H} , let $P(E|H_i) = kP(E'|H_i)$. Then from Bayes' theorem, we have:

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

Given the likelihood theorem, we can derive the LP if we make the following assumption about quantitative confirmation:

(C) If $P(H|E) = P(H|E')$, then $c(H, E) = c(H, E')$.

Ward Edwards et al. (1963, p. 327) present essentially this argument for the LP: a Bayesian should accept the LP because it follows from the likelihood theorem and (C). However, they do not explicitly formulate (C), and they give no indication of being aware that it is a substantive assumption required for their argument. Indeed, Bayesian analyses of various aspects of scientific methodology often assume implicitly (C) and its companion $c(H, E) > c(H, E')$ if $P(H|E) > P(H|E')$ (cf. Horwich 1982, Maher 1988). Furthermore, (C) is entailed by the confirmation measures d , r , and l . Recall that $d(H, E) =_{\text{df}} P(H|E) - P(H)$, while

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

It is trivial that (C) holds when $c(H, E)$ is d or r , so I provide the proof only in the case in which $c(H, E) = l(H, E)$. Recall that

$$l(H, E) =_{\text{df}} \log \left[\frac{P(E|H)}{P(E|\bar{H})} \right].$$

Suppose that $P(H|E) = P(H|E')$. Then from Bayes' theorem we have:

$$\frac{P(E|H)}{P(E|\bar{H})} = \frac{P(H|E)P(\bar{H})}{P(\bar{H}|E)P(H)} = \frac{P(H|E')P(\bar{H})}{P(\bar{H}|E')P(H)} = \frac{P(E'|H)}{P(E'|\bar{H})}$$

Hence, if $P(H|E) = P(H|E')$, then $l(H, E) = l(H, E')$. A Bayesian who accepts d , r , or l as her confirmation measure, therefore, is committed to the LP.

Since (C) together with a theorem of the probability calculus entails the LP, it follows that any Bayesian confirmation measure that violates the LP must also violate (C). And indeed, it is easy to see that any k -measure does so. That is, if the antecedent of the likelihood theorem holds and μ is a k -measure, then for all H_i in \mathbf{H} , $P(H_i|E) = P(H_i|E')$ while $\mu(H_i, E) = k\mu(H_i, E')$. Moreover, we can show that ρ , n , and m are k -measures.

Since we have seen already that ρ , n , and m are Bayesian confirmation measures, it only remains to show that these measures satisfy article (b) of the definition of a k -measure. Let us begin with ρ . Suppose that, for all $H_i \in \mathbf{H}$, $P(E|H_i) = kP(E'|H_i)$, where $k > 0$. Then we have:

$$\begin{aligned} (1) \quad P(E) &= \sum_i P(H_i)P(E|H_i) = \sum_i P(H_i)kP(E'|H_i) \\ &= k \sum_i P(H_i)P(E'|H_i) = kP(E'). \end{aligned}$$

Let H be an arbitrarily chosen member of \mathbf{H} . But now:

$$\begin{aligned}
 (2) \quad \rho(H, E) &=_{\text{df}} P(H \& E) - P(H)P(E) \\
 &= P(H)P(E|H) - P(H)P(E) \\
 &= P(H)kP(E'|H) - P(H)kP(E') \\
 &= k[P(H)P(E'|H) - P(H)P(E')] \\
 &= k[P(H \& E') - P(H)P(E')] \\
 &= k\rho(H, E')
 \end{aligned}$$

Therefore, if $P(E|H_i) = kP(E'|H_i)$ for each $H_i \in \mathbf{H}$, then $\rho(H_i, E) = k\rho(H_i, E')$ for each $H_i \in \mathbf{H}$.

The proof in the case of m is similar. Suppose that $P(E|H_i) = kP(E'|H_i)$ for each $H_i \in \mathbf{H}$. Let H be an arbitrarily chosen member of \mathbf{H} . Then from (1) we have:

$$\begin{aligned}
 (3) \quad m(H, E) &=_{\text{df}} P(E|H) - P(E) = k(E'|H) - kP(E') \\
 &= k[P(E'|H) - P(E')] = km(H, E').
 \end{aligned}$$

The proof in the case of n follows easily, once we note that, by the theorem of total probability,

$$P(E|\bar{H}_a) = \sum_{i \neq a} P(H_i|\bar{H}_a)P(E|H_i),$$

since for every $i \neq a$, H_i entails \bar{H}_a . Suppose that $P(E|H_i) = kP(E'|H_i)$ for each $H_i \in \mathbf{H}$, and let H_a be an arbitrarily chosen member of \mathbf{H} . Then we have:

$$\begin{aligned}
 (4) \quad n(H_a, E) &=_{\text{df}} P(E|H_a) - P(E|\bar{H}_a) \\
 &= P(E|H_a) - \sum_{i \neq a} P(H_i|\bar{H}_a)P(E|H_i) \\
 &= kP(E'|H_a) - \sum_{i \neq a} P(H_i|\bar{H}_a)kP(E'|H_i) \\
 &= k \left[P(E'|H_a) - \sum_{i \neq a} P(H_i|\bar{H}_a)P(E'|H_i) \right] \\
 &= k[P(E'|H_a) - P(E'|\bar{H}_a)] = kn(H_a, E').
 \end{aligned}$$

Thus, ρ , n , and m are k -measures.

Let us consider how these observations relate to the example of the binomial and negative binomial experiments given above in Section 2. In that example, we considered two experiments. In the first (the binomial), it was decided that the coin would be flipped ten times and the number of heads recorded. The outcome was five heads. In the second (the negative binomial), we decided that we would keep flipping the coin until we got five heads. The outcome was that the fifth head was obtained on the tenth flip. Let E describe the outcome of the binomial experiment and E' the outcome of the negative binomial experiment. Suppose that the H_i 's in \mathbf{H} each specify a distinct probability that the coin will come up heads when flipped. Then we saw that $P(E|H_i) = 2P(E'|H_i)$, for each $H_i \in \mathbf{H}$. Thus, since ρ is a k -measure, $\rho(H_i, E) = 2\rho(H_i, E')$. Hence, if we accepted ρ , we would judge in this case that E provides greater confirmation (or less disconfirmation) for H than E' .

Thus, a Bayesian who took ρ , m , or n as his confirmation measure would reject the LP. Such a Bayesian would also say that the stopping rule is relevant to the evidential assessment of an experimental result. However, making $c(H, E)$ a k -measure comes at a cost, since Bayesian analyses that assume (C) in the process of accounting for some feature of scientific methodology would have to be dismissed as unsound. A further difficulty is that if $c(H, E)$ is a k -measure, a very useful principle of Bayesian statistics must be rejected. It is to this issue that I now turn.

5. THE SUFFICIENCY PRINCIPLE

The statistician Allan Birnbaum (1962) showed that the LP could be derived from two suppositions, namely, the principles of *sufficiency* and *conditionality*. The sufficiency principle asserts something about what information can be omitted in the description of an experimental outcome without altering its evidential force. This principle is commonly accepted by Bayesian and frequentist statisticians alike. The conditionality principle, stated roughly, asserts that “experiments not actually performed should be irrelevant to conclusions” (Berger and Wolpert 1988, p. 1). Since the LP follows from the sufficiency and conditionality principles, one who rejects the LP must also reject at least one of them. In this section, I show that a Bayesian who rejects the LP will also be strongly inclined to reject the sufficiency principle. In particular, if $c(H, E)$ is a k -measure, then the sufficiency principle no longer holds.

The sufficiency principle asserts something about what sorts of statistics are appropriate for characterizing the data produced by an experiment. For example, in the coin flipping case, one could provide the complete list

of outcomes of ordered throws of the coin, or one might record the number of heads, or one might only record the ratio of heads to tails. Each of these different ways of describing the outcome of the experiment is a different statistic and, as the example shows, some statistics contain more information than others. Given that more work is required to record statistics that provide more detailed information, it is natural to desire that our statistic include all the relevant information and no more. The sufficiency principle addresses exactly this point.

The expression *sufficient statistic* is defined as follows:

Sample statistic t is said to be sufficient, relative to the parameter of interest, θ , if the probability of any particular member of the outcome space given t , is independent of θ . (Howson and Urbach 1993, p. 189)

Suppose we let \mathbf{x} represent the outcome space. Thus, in our example in which we flip a coin ten times, the outcome space would be the set of all possible ordered combinations of heads and tails in ten flips of a coin. The statistic t , then, is a function that maps members of \mathbf{x} onto a set of numbers. For example, t could be defined as *the number of heads*. Suppose that each member of \mathbf{H} assigns a value to the parameter θ . For example, θ could represent the probability that the coin will come up heads when flipped. Then we can state the definition of a sufficient statistic in Bayesian terms as follows:

t is a sufficient statistic with respect to θ if and only if, for all $H_i \in \mathbf{H}$, for all $x \in \mathbf{x}$, and all values v of t , $P(x|v \ \& \ H_i) = P(x|v)$.

For convenience, the clause to the right of the “if and only if” is commonly abbreviated as $P(\mathbf{x}|t \ \& \ \theta) = P(\mathbf{x}|t)$. I shall extend this notational shortcut as follows: $c(\theta, t) = c(\theta, \mathbf{x})$ is equivalent to, for each $H_i \in \mathbf{H}$, $c(H_i, t) = c(H_i, \mathbf{x})$.

The *sufficiency principle*, then, tells us that a sufficient statistic omits no information about the experimental outcome that is relevant to assessing its evidential force (cf. Birnbaum 1962, p. 270). The sufficiency principle is mainly useful as a justification for various labor saving procedures, since it assures us that certain types of information can be omitted without any cognitive loss. Of course, this does not mean that no information besides that included in the sufficient statistic is relevant; rather, it means that any information *about the outcome* not provided by the sufficient statistic is evidentially irrelevant. In our formalism, the sufficiency principle can be put as follows:

SUFFICIENCY PRINCIPLE. If t is a sufficient statistic with respect to θ , then $c(\theta, t) = c(\theta, \mathbf{x})$.

That is, the principle states that if t is a sufficient statistic, then describing the outcome in further detail makes no difference to the degree of confirmation conferred upon the hypothesis.

As Howson and Urbach show (1993, 227), it is easy to give a Bayesian argument for the sufficiency principle. The argument proceeds as follows. Suppose that t is a sufficient statistic with respect to θ . Now by Bayes' theorem, we have:

$$(1) \quad P(\mathbf{x}|\theta \& t) = P(\mathbf{x}|t) \times [P(\theta|\mathbf{x} \& t)/P(\theta|t)]$$

Since the value of t is determined by \mathbf{x} , (1) simplifies to:

$$(2) \quad P(\mathbf{x}|\theta \& t) = P(\mathbf{x}|t) \times [P(\theta|\mathbf{x})/P(\theta|t)]$$

But since t is a sufficient statistic with respect to θ , $P(\mathbf{x}|\theta \& t) = P(\mathbf{x}|t)$. Hence, from (2) we have:

$$(3) \quad P(\mathbf{x}|t) = P(\mathbf{x}|t) \times [P(\theta|\mathbf{x})/P(\theta|t)]$$

From (3) it follows that:

$$(4) \quad P(\theta|\mathbf{x}) = P(\theta|t)$$

Howson and Urbach's argument for the sufficiency principle stops at this point. But no doubt the following step is implicit. From (4), (C) enables us to conclude that:

$$(5) \quad c(\theta, \mathbf{x}) = c(\theta, t)$$

Thus, if (C) is granted, then we have a straightforward Bayesian derivation of the sufficiency principle.

However, we saw in the foregoing section that relinquishing the LP necessitates abandoning (C). So a Bayesian who rejected the LP would have to regard the above argument for the sufficiency principle as unsound. Indeed, it is easy to invent examples in which the sufficiency principle is false if $c(H, E)$ is a k -measure. For instance, consider a simple binomial example in which θ is the probability that a particular coin will come up heads when flipped. Suppose that each member of \mathbf{H} ascribes a distinct value to θ . Let our experiment consist of two flips of the coin, so the outcome space is as follows: $\mathbf{x} = \{HH, HT, TH, TT\}$, where "H" and

“ T ” are interpreted in the obvious way. Let the statistic t represent the number of heads. For each $H_i \in \mathbf{H}$,

$$\begin{aligned}
 (6) \quad P(HH|H_i \ \& \ t = 2) &= P(HH|t = 2) = 1 \\
 P(HT|H_i \ \& \ t = 1) &= P(HT|t = 1) = P(TH|H_i \ \& \ t = 1) \\
 &= P(TH|t = 1) = 0.5 \\
 P(TT|H_i \ \& \ t = 0) &= P(TT|t = 0) = 1.
 \end{aligned}$$

Thus, we have that $P(\mathbf{x}|\theta \ \& \ t) = P(\mathbf{x}|t)$, and hence t is a sufficient statistic with respect to θ , so the antecedent of the sufficiency principle is satisfied.

For each $H_i \in \mathbf{H}$, let p_i be the value that H_i ascribes to θ . Then by the binomial formula, we have for each $H_i \in \mathbf{H}$:

$$\begin{aligned}
 (7) \quad P(HT|H_i) &= p_i(1 - p_i) \\
 (8) \quad P(t = 1|H_i) &= 2p_i(1 - p_i).
 \end{aligned}$$

Thus, from (7) and (8) we have:

$$(9) \quad P(t = 1|H_i) = 2P(HT|H_i), \text{ for each } H_i \in \mathbf{H}.$$

Thus, if $c(H, E)$ is a k -measure, $c(H_i, t = 1) = 2c(H_i, HT)$, for each $H_i \in \mathbf{H}$, which contradicts the sufficiency principle.

6. CONCLUSION

One can be a Bayesian and yet accept that stopping rules are evidentially relevant by choosing a Bayesian confirmation measure that is also a k -measure, of which several have been seriously proposed. All k -measures violate the LP and hence enable the stopping rule to influence confirmation. However, two major inconveniences ensue from adopting a k -measure as one's Bayesian measure of confirmation. First, the common Bayesian assumption that E and E' confirm H equally if $P(H|E) = P(H|E')$ must be given up. Second, the sufficiency principle, which is of practical use in statistics, must also be abandoned. The desire to retain (C) and the sufficiency principle are sensible reasons for not choosing a k -measure as one's Bayesian measure of confirmation, which thereby provides an indirect Bayesian argument for the LP. However, it nevertheless remains the case that one could be a Bayesian, while rejecting the LP and accepting the evidential relevance of stopping rules. Therefore,

unconditional claims to the effect that Bayes' theorem entails the LP are oversimplifications of a much more complex situation.

NOTES

* I would like to thank two anonymous referees for helpful comments on an earlier draft of this essay.

¹ This objection seems to have been first made by the statistician P. Armitage in the course of a discussion at a statistics conference (Savage 1962, 72). The argument is also discussed by Berger and Wolpert (1988, pp. 80–82).

² See Birnbaum (1962, p. 271), Edwards, Lindman and Savage (1963, pp. 237–240), and Berger and Wolpert (1988, p. 19).

³ For a more general justification and discussion of this claim, see Birnbaum (1962) and Berger and Wolpert (1988, pp. 74–79).

⁴ Some Bayesians would accept these principles only with qualifications. For example, several Bayesians have argued that although it is generally reasonable to assume strict conditionalization, this principle may fail under certain circumstances (cf. Maher 1993; Howson 1997). Moreover, objective Bayesians would restrict the principles of qualitative confirmation to cases in which P satisfies additional constraints of rationality beyond the axioms of probability (cf. Maher 1996). However, such qualifications make no difference to the argument presented here.

⁵ Andrew Backe also asserts that Bayes' theorem entails the likelihood principle (1999, p. 354).

⁶ A version of ρ was proposed by Carnap (1962, 360), while n and m have been proposed by Robert Nozick (1981, p. 252) and H. Mortimer (1988, §11.1), respectively.

⁷ Fitelson (2001, S125–S130) argues in favor of l .

REFERENCES

- Backe, A.: 1999, 'The Likelihood Principle and the Reliability of Experiments', *Philosophy of Science* **66**, S354–S361.
- Berger, J. and R. Wolpert: 1988, *The Likelihood Principle*, 2nd ed, Institute of Mathematical Statistics, Hayward, CA.
- Birnbaum, A.: 1962, 'On the Foundations of Statistical Inference', *Journal of the American Statistical Association* **57**, 269–326.
- Carnap, R.: 1962, *Logical Foundations of Probability*, University of Chicago Press, Chicago.
- 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.
- Fitelson, Branden: 1999, 'The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity', *Philosophy of Science* **66**, S362–S378.
- Fitelson, Branden: 2001, 'A Bayesian Account of Independent Evidence with Applications', *Philosophy of Science* **68**, S123–S140.
- Horwich, P.: 1982, *Probability and Evidence*, Cambridge University Press, Cambridge.

- Howson, Colin: 1997, 'A Logic of Induction', *Philosophy of Science* **64**, 268–290.
- Howson, C. and Urbach, P.: 1993, *Scientific Reasoning: the Bayesian Approach*, 2nd edn, Open Court, La Salle, IL.
- 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.
- Kyburg, H.: 1974, *The Logical Foundations of Statistical Inference*, Dordrecht: D. Reidel Publishing Co.
- Maher, P.: 1988, 'Prediction, Accommodation, and the Logic of Discovery', in A. Fine and J. Leplin (eds.) *PSA 1988, Vol. 1*, Philosophy of Science Association, East Lansing, MI, pp. 273–285.
- 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.
- Mayo, D: 1996, *Error and the Growth of Experimental Knowledge*, Chicago University Press, Chicago IL.
- Milne, P.: 1996, ' $\log[p(h/eb)/p(h/b)]$ is the One True Measure of Confirmation', *Philosophy of Science* **63**, 21–26.
- Mortimer, H.: 1988, *The Logic of Induction*, Prentice Hall, Paramus.
- Nozick, R.: 1981, *Philosophical Explanations*, Harvard University Press, Cambridge, MA.
- Pratt, J.: 1962, In A. Birnbaum (ed.), 'On the Foundations of Statistical Inference', *Journal of the American Statistical Association* **57**, 269–326.
- Savage, L.: 1962, *Foundations of Statistical Inference: A Discussion*, Methuen, London.
- Schlesinger, G.: 1995, 'Measuring Degrees of Confirmation', *Analysis* **55**, 208–211.
- Williamson, J.: 1999, 'Countable Additivity and Subjective Probability', *British Journal for the Philosophy of Science* **50**, 401–416.

Department of Philosophy
 Michigan State University
 503 S. Kedzie Hall
 East Lansing, MI 48824-1032
 USA
 E-mail: dpsst19@yahoo.com

Manuscript submitted 19 December, 2001
 Final version received 29 May, 2002

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.