



Bayesianism and Diverse Evidence

Andrew Wayne

Philosophy of Science, Vol. 62, No. 1. (Mar., 1995), pp. 111-121.

Stable URL:

<http://links.jstor.org/sici?sici=0031-8248%28199503%2962%3A1%3C111%3ABADE%3E2.0.CO%3B2-G>

Philosophy of Science is currently published by The University of Chicago Press.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ucpress.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

BAYESIANISM AND DIVERSE EVIDENCE*

ANDREW WAYNE^{†‡}

*Department of Philosophy
University of Rochester*

A common methodological adage holds that diverse evidence better confirms a hypothesis than does the same amount of similar evidence. Proponents of Bayesian approaches to scientific reasoning such as Horwich, Howson and Urbach, and Earman claim to offer both a precise rendering of this maxim in probabilistic terms and an explanation of why the maxim should be part of the methodological canon of good science. This paper contends that these claims are mistaken and that, at best, Bayesian accounts of diverse evidence are crucially incomplete. This failure should lend renewed force to a long-neglected global worry about Bayesian approaches.

1. Introduction. Bayesian approaches to understanding scientific reasoning have proven extremely resilient in the face of extensive criticism. A number of critics, for instance, have raised compelling concerns about the justifiability of the practice, central to all Bayesian approaches, of assigning point-valued probability functions to degrees of belief of rational agents. H. Kyburg (1978) has argued that point-valued probability assignments impose an illegitimate and unrealistic precision upon the description of an agent's belief commitments. The actual indefiniteness in an agent's beliefs, he contends, is better described by interval-valued probability functions where the size of the interval reflects the agent's uncertainty in degree of belief. An interval-valued probability calculus is much weaker than a point-valued calculus, and many of the relations between prior and posterior probabilities on which Bayesian approaches depend no longer hold. So much the worse for Bayesian approaches to scientific reasoning, critics conclude.

Curiously, despite a lack of cogent replies to these worries, they have had remarkably little effect on Bayesians. Bayesians invariably agree that Bayesianism is in many ways unrealistic and is best regarded as articulating an idealization of scientific inference, one which is satisfied to a greater or lesser degree in actual scientific method. They argue that the

*Received July 1993; revised April 1994.

[†]I would like to thank Philip Kitcher, Paul Teller, Josh Jorgensen and an anonymous referee for incisive comments on earlier drafts of this paper. This work was supported by a Fellowship from the Social Sciences and Humanities Research Council of Canada.

[‡]Send reprint requests to the author, Department of Philosophy, University of Rochester, Rochester, NY 14627, USA.

Philosophy of Science, 62 (1995) pp. 111–121

Copyright © 1995 by the Philosophy of Science Association.

incremental gain in fidelity to practice by adopting interval-valued probability assignments is far outweighed by the consequent loss of explanatory power. Bayesians typically close their discussions of this problem as follows:

Let us regard the issue of point versus interval-valued probabilities as now sufficiently discussed and take the proof of the pudding to be in the eating: we feel confident that by the end of the book the reader will have become convinced of the utility of the point-probability model. (Howson and Urbach 1989, 70)

By the end of this exposition of Bayesian methods the reader has been treated to a plethora of ostensible explanatory and practical successes. The moral one is expected to come away with is that Bayesians may make some idealizing assumptions, but they have been successful at rendering precise and explaining a wide range of specific elements of good scientific reasoning. It is not methodologically sound to reject a dynamic and practically successful research program on the basis of skepticism about some of its idealizing assumptions.

If Bayesian methods were shown to be far less successful than advertised at rendering precise and explaining particular elements of good scientific method, then Bayesians' "proof in the pudding" defense would be far less plausible and the global worries sketched above would gain renewed force. The present paper takes up the project of establishing the antecedent by examining Bayesian accounts of an important element in good scientific method: the notion of *diverse evidence*. A common methodological adage holds that diverse evidence better confirms a hypothesis than does the same amount of similar evidence. Bayesians claim to offer both a precise rendering of this maxim in probabilistic terms and an explanation of why the maxim should be part of the methodological canon of good science—one recent book goes so far as to place the two major Bayesian accounts of diverse evidence in a chapter boldly titled "Success Stories" (Earman 1992). This paper contends that, at best, Bayesian accounts of diverse evidence are crucially incomplete, and that reports of their success are greatly exaggerated.

2. The Correlation Approach. The simplest and most common Bayesian approach to diverse evidence relates the diversity of a data set to the degree of probabilistic independence among its members. Recent proponents of this view include A. Franklin and C. Howson (1984), P. Urbach (Howson and Urbach 1989), and J. Earman (1992). The idea is that if one result, say from a measurement of the acceleration due to gravity performed last Monday, raises our estimation of the likelihood of a second result, for example the same result for the same experiment

performed next Monday, then these two pieces of evidence are clearly similar. Conversely, the result of a measurement of the acceleration due to gravity performed last Monday gives no reason to raise our expectation of the likelihood of evidence for life on Mars; these pieces of evidence are diverse.

Howson and Urbach capture this idea in Bayesian terms for the simple case in which a data set has two elements, $e = \{e_1, e_2\}$:

The idea of similarity between items of evidence is expressed naturally in probabilistic terms by saying that e_1 and e_2 are similar if $\Pr(e_2/e_1)$ is higher than $\Pr(e_2)$ and one might add that the more the first probability exceeds the second, the greater the similarity. (1989, 114; my notation)

(We assume an implicit conditionalization on background assumptions, K . Thus $\Pr(e_2/e_1) = \Pr(e_2/e_1 \ \& \ K)$.) I propose a measure of similarity $S(e_1, e_2)$ between two pieces of evidence e_1 and e_2 such that

$$\Pr(e_2/e_1) =_{df} \Pr(e_2)S(e_1, e_2). \tag{1}$$

When $S(e_1, e_2)$ is much greater than one, on Howson and Urbach’s account, e_1 and e_2 are similar items of evidence, and the data set $\{e_1, e_2\}$ is called similar or narrow. When $S(e_1, e_2)$ is close to one, e_1 and e_2 are said to constitute a diverse data set. Using the definition of conditional probability from elementary probability calculus,

$$\text{for } \Pr(b) > 0, \quad \Pr(a/b) =_{df} \Pr(a \& b) / \Pr(b),$$

(1) may be rewritten:

$$\Pr(e_2 \& e_1) = \Pr(e_1)\Pr(e_2)S(e_1, e_2). \tag{2}$$

In the case of a diverse data set, $S(e_1, e_2) \approx 1$, (2) is the familiar statement of probabilistic independence between two items of evidence. In the case of a narrow data set, $S(e_1, e_2) \gg 1$, (2) states that a high degree of correlation exists between the items of evidence. Thus Howson and Urbach’s (ibid., 113) Bayesian explication of diversity among items of evidence is not simply, as they say, “closely related to” the ideas of probabilistic correlation and independence; rather, the two notions are identical. For Howson and Urbach, degree of diversity simply *is* degree of probabilistic independence.

Earman (1992, 78) has recently extended this approach to the general case in which a data set contains more than two elements. Earman defines diversity of evidence in terms of “the rate of increase in the factors $\Pr(e_k/e_1 \& e_2 \& \dots \& e_{k-1})$ ”; the slower the rate of increase, the more diverse the data set. The superior confirmatory value of diverse evidence is then

accounted for as follows. Consider Bayes's theorem for cases in which h entails e . The theorem may then be written in the form:

$$\Pr(h/e) = \Pr(h)/\Pr(e_1)\Pr(e_2/e_1) \dots \Pr(e_k/e_1\&e_2\& \dots \&e_{k-1}), \quad (3)$$

where $e = \{e_1, e_2, \dots, e_k\}$. Now for a given k , Earman reasons, the more slowly the factors $\Pr(e_i/e_1\&e_2\& \dots \&e_{i-1})$ increase, the smaller will be the denominator on the right-hand side. For a given k , the smaller the denominator, the larger the posterior probability of h . By definition, diverse evidence is that for which the factors in the denominator tend to increase slowly. Thus we see how diverse evidence is able to boost the posterior probability of an hypothesis more effectively than does similar evidence.

Earman leaves us with the incorrect impression that diverse evidence always confirms better than does the same amount of similar evidence. To see the problem, think back to the weekend of 9 and 10 November 1974, when a new particle, called the J/ψ , was independently discovered at two particle accelerators. This unexpected discovery, later dubbed "the November revolution" by high-energy physicists, provided strong support for the electro-weak model of subatomic interactions. The discovery, in fact, provided much stronger support than did a wide range of far less surprising results—about such things as particle cross sections and decay times—which are also consistent with the electro-weak model. In this case the sheer unexpectedness of the discoveries, that is, their very low prior probabilities, offsets their high correlation. Intuitively, sometimes an unexpected pair of results which are highly correlated can boost the probability of an hypothesis more than a pair of diverse results with relatively high priors.

In his analysis, Earman fails to distinguish two distinct components in the denominator of (3) which contribute to how fast it will grow. The rate of increase of factors $\Pr(e_i/e_1\&e_2\& \dots \&e_{i-1})$ depends not only on the degree of correlation among items of evidence *but also depends on their individual prior probabilities*. To see this most easily, we need to generalize the measure of similarity defined in (1) to data sets with an arbitrary number of members. For a given item of evidence e_k , $S(e_1, e_2, \dots, e_k)$ (hereafter abbreviated S_k) is defined as a measure of degree of similarity (or probabilistic correlation) between e_k and the other items in the set:

$$\Pr(e_k)S_k =_{\text{def}} \Pr(e_k/e_1\&e_2\& \dots \&e_{k-1}). \quad (4)$$

Earman's definition of diversity of evidence can be restated as saying that a data set is diverse to the extent that the rate of increase of the factors $\Pr(e_i)S_i$ is generally slow, while a data set is similar to the extent that the

rate of increase of these factors is extreme. Combining (3) and (4), Bayes's theorem yields

$$\Pr(h/e) = \Pr(h)/\Pr(e_1)\Pr(e_2)S_2 \dots \Pr(e_k)S_k. \quad (5)$$

Equation (5) manifests the two contributions to the rate of increase of the factors $\Pr(e_i/e_1 \& e_2 \& \dots \& e_{i-1})$: the individual priors (the $\Pr(e_i)$) and the degree of probabilistic correlation among them (the S_i).

Conflating these two contributions vitiates Earman's definition of diverse evidence. The correlation approach attempts to capture in Bayesian terms an intuition which says roughly that diversity has to do with how our initial degrees of belief in items of evidence are modified as we conditionalize on other evidence. Thus the diversity of a data set cannot be judged, as Earman asserts, by the rate of increase of the factors $\Pr(e_i/e_1 \& e_2 \& \dots \& e_{i-1})$, for this rate of increase depends on the individual priors in a highly counterintuitive way.

The problem is easily remedied. Let me propose the following definition of diversity, which captures the central intuition of the correlation approach: *A data set is similar (diverse) to the extent that the rate of increase of the factors S_i is extreme (slow)*. More formally, this is most clearly expressed in terms of a comparative condition on rival data sets:

(A) For $e = \{e_1, e_2, \dots, e_k\}$ with measures of similarity S_i and rival $e' = \{e'_1, e'_2, \dots, e'_k\}$ with measures of similarity S'_i , e is a significantly more diverse data set than e' just in case the members of e are significantly less correlated than the members of e' , that is, $S(e_1, e_2)$ is smaller than $S(e_1, e_2)'$ and the other S_i increase much more slowly than the S'_i .

Condition (A) does not apply when the rate of increase is equivocal between S_i and S'_i ; in this case it is unable to make a reliable discrimination. The vagueness in (A) reflects the limits of applicability of the comparative notion of diverse evidence as it is used in best scientific practice. An important consequence of (A) is that diverse evidence confirms better than does the same amount of similar evidence when the individual priors are comparable.

Do we, as Howson and Urbach assert, have a Bayesian explanation of the value of diverse evidence? Is this an instance, as Earman claims, of "providing a Bayesian rationale for what are regarded as sound methodological procedures" (ibid., 63)? I think not. It was sound, before Maxwell, to regard electromagnetic and optical phenomena as diverse, while today they are judged very similar. Clearly, judgements of diversity are always made relative to a given theoretical context; what look like disparate phenomena in one context may appear closely akin in another. Without a satisfactory Bayesian account of how judgements of diver-

sity—assignments of degree of correlation among elements of a data set—depend on the theoretical context, the correlation approach is crucially incomplete.

Yet Bayesianism is always incomplete—it does not purport to explain our prior probability assignments, for instance. And this is not taken to impair its ability to explain and justify many elements of good scientific reasoning (at least so long as questions about the overarching rationality or objectivity of Bayesian approaches are neglected). Bayesians often gloss this by saying that Bayesian explanations of specific elements of good scientific reasoning take as their starting point an assignment of values to the priors. My objection is not simply that the correlation approach is incomplete in this way, but that it suffers a further lack. Our understanding of variety of evidence involves intuitions about how judgements of diversity are affected by the theoretical context, intuitions which underwrite, for example, our understanding of the case of electromagnetic and optical evidence before and after Maxwell. The correlation approach remains crucially incomplete until it is supplemented by a Bayesian account of these intuitions. If this were done successfully, I suggest, it would constitute a large step toward counting diverse evidence among Bayesian success stories.

3. The Eliminative Approach. Fortunately, a well-known Bayesian analysis of diverse evidence due to P. Horwich (1982, 118–122) attempts to do precisely this. Horwich begins with the intuition that “diverse data tend to eliminate from consideration many of the initially most plausible, competing hypotheses. Narrow data, on the other hand, leave many initially very plausible alternatives in the field, and can therefore provide relatively little reason to select any one of them” (*ibid.*, 118). Horwich purports to explain, in Bayesian terms, why we judge certain data to be diverse: Diverse data are more efficient at eliminating large chunks of the space of plausible rival hypotheses than are similar data.

Horwich offers a comparative condition to judge evidential variety. Consider an hypothesis under test h_1 and a set of exhaustive and mutually exclusive alternative hypotheses $\{h_1, h_2, \dots, h_k\}$, one of which is correct. Given two data sets, e and e' :

(B) e is a more diverse data set than e' iff, for many of the alternative hypotheses h_j with substantial prior probabilities,

$$\Pr(e/h_j) < \Pr(e'/h_j),$$

(adapted from Horwich [*ibid.*, 119]; implicit in this condition is the rider that in order for e to be judged more diverse there must be few if any cases of the converse situation in which $\Pr(e/h_j)$ is greater than $\Pr(e'/h_j)$).

From this condition, claims Horwich, a Bayesian explanation of the superior evidential value of varied evidence follows. Given our assumptions about the h_i , the prior probability of an item of evidence e can be written

$$\Pr(e) = \Pr(h_1)\Pr(e/h_1) + \Pr(h_2)\Pr(e/h_2) + \dots + \Pr(h_k)\Pr(e/h_k). \quad (6)$$

Horwich (ibid.) assumes that both e and e' are entailed by h_1 , and uses Bayes's theorem to write the ratio of the posterior probability of h_1 given e to h_1 given e' as

$$\begin{aligned} \Pr(h_1/e)/\Pr(h_1/e') &= [\Pr(h_1) + \Pr(h_2)\Pr(e'/h_2) + \dots \\ &+ \Pr(h_k)\Pr(e'/h_k)]/[\Pr(h_1) + \Pr(h_2)\Pr(e/h_2) + \dots \\ &+ \Pr(h_k)\Pr(e/h_k)]. \end{aligned} \quad (7)$$

Given **(B)**, (7) implies that $\Pr(h_1/e) > \Pr(h_1/e')$ and the diverse evidence lends a greater confirmational boost than does the similar evidence.

If successful, this account of diverse evidence would serve to underwrite and explain hitherto brute facts about correlations within a data set. It would show that we assign a low (high) degree of correlation to that data set which is (is not) efficient at eliminating plausible alternative hypotheses. To see how this might work in a simple situation, imagine that e and e' each contain two items of evidence: $e = \{e_1, e_2\}$ and $e' = \{e'_1, e'_2\}$. For simplicity assume that prior to conditionalizing on any h_j the degree of correlation is the same for e and e' , and all four elements of e and e' (e_1, e_2, e'_1 , and e'_2) have the same individual probabilities both prior to and after conditionalizing on h_j . Now examine what happens when we conditionalize on h_j , assuming this to be one of the "many alternative hypotheses" stipulated by **(B)**. According to **(B)**, e is more efficient at eliminating rival hypotheses than e' just in case $\Pr(e/h_j) < \Pr(e'/h_j)$, which in this case can be written

$$\Pr(e_1 \& e_2/h_j) < \Pr(e'_1 \& e'_2/h_j). \quad (8)$$

From (8), the above assumptions, and definition (1), a simple calculation shows that *the effect of conditionalizing on h_j is to render $S(e_1, e_2) < S(e'_1, e'_2)$* . Thus in this simple case, at least, degree of correlation is a direct consequence of efficacy at eliminating rival hypotheses. This is admittedly a contrived example of how judgements of degree of correlation are dependent on the range of alternative hypotheses, but it makes plausible the claim that if Horwich's account were successful some such Bayesian account could be developed.

However, Horwich's account is not successful, for it contains a serious inconsistency. Horwich states that h_1 entails the data ($\Pr(e/h_1) = 1$); h_1 is thus a *deterministic hypothesis* in the sense that it makes determinate predictions about evidential states. Yet the other h_j , according to Horwich, do not entail the data ($\Pr(e/h_j) < 1$); they are *statistical hypotheses* which

make exclusively probabilistic assertions about evidential states. Traditionally philosophers of science have been concerned with constructing Bayesian accounts of scientific reasoning which deal with deterministic hypotheses (often labelled “the hypothetico-deductive case”), while the details of Bayesian accounts involving statistical hypotheses have been left largely to statisticians. The important point for our purposes is that in the vast majority of scientific circumstances the hypothesis under test and its rivals *will be of the same kind*. Horwich presupposes—in (7)—a case which is rarely (if ever) instantiated in science in which the hypothesis under test is deterministic and all the alternatives are statistical. What becomes of Horwich’s analysis when this assumption is removed? Consider the deterministic and statistical cases in turn.

For all hypotheses to be deterministic means that an item of evidence is either entailed by or inconsistent with each hypothesis, $\Pr(e/h_j) = 1$ or 0. Horwich’s condition (**B**) becomes:

(**B**_{det}) *e* is a more diverse data set than *e'* iff, for many of the alternative hypotheses h_j with substantial prior probabilities, $\Pr(e/h_j) = 0$ while $\Pr(e'/h_j) = 1$.

(As with (**B**), the implicit rider here is that there have to be few cases in which $\Pr(e/h_j) = 1$ while $\Pr(e'/h_j) = 0$.) In this case (7) reduces to

$$[\Pr(h_1/e)/\Pr(h_1/e')] = [\sum_j \Pr(h_i)/\sum_j \Pr(h_j)], \quad (9)$$

where $\{h_i\}$ is the set of all those hypotheses not eliminated by the similar evidence and $\{h_j\}$ is the set of all those hypotheses not eliminated by the diverse evidence. By condition (**B**_{det}), $\{h_i\}$ is larger than $\{h_j\}$, and from (9) the desired result is obtained: The diverse evidence supports h_1 more than does the similar evidence.

As an explanation of the superior evidential value of a diverse data set this account is clearly circular. Diverse evidence is better, the explanation goes, because of its ability to eliminate more of the rival hypotheses, yet eliminating more of the rival hypotheses is exactly the definition of diverse evidence with which Horwich began. Such a circularity is tolerable if the resulting account yields a deeper understanding of the notion of diverse evidence. But in this case it does not, for the notion of diverse evidence, condition (**B**_{det}), and the explanation of (**B**_{det})’s efficacy are all three simply restatements, using symbols of probabilistic algebra, of Horwich’s original intuition.

Perhaps Bayesianism has nothing revealing to say about the superior value of diverse evidence in a straightforwardly deterministic context. In any case, Horwich is primarily concerned with something closer to a statistical context. It is plausible that Horwich meant to discuss the value of

diverse evidence in statistical contexts but simply made a slip when he asserted h_1 to be a deterministic hypothesis.

Let us see if sense can be made of a Bayesian explanation in a uniformly statistical context. Now $\Pr(e/h_1) < 1$ and (7), the ratio of the posterior probability of the hypothesis under test given diverse evidence e and narrow evidence e' , becomes

$$\begin{aligned} \Pr(h_1/e)/\Pr(h_1/e') &= \Pr(e/h_1)[\Pr(h_1)\Pr(e'/h_1) \\ &+ \Pr(h_2)\Pr(e'/h_2) + \dots \\ &+ \Pr(h_k)\Pr(e'/h_k)]/\Pr(e'/h_1) [\Pr(h_1)\Pr(e/h_1) + \Pr(h_2)\Pr(e/h_2) \\ &+ \dots + \Pr(h_k)\Pr(e/h_k)]. \end{aligned} \tag{10}$$

For the diverse evidence to confirm better the ratio in (10) must be larger than one. The ratio of the right-hand side of (10) is larger than one just in case

$$\begin{aligned} \Pr(e/h_1)/\Pr(e'/h_1) &> [\Pr(h_1)\Pr(e/h_1) + \Pr(h_2)\Pr(e/h_2) + \dots \\ &+ \Pr(h_k)\Pr(e/h_k)]/[\Pr(h_1)\Pr(e'/h_1) + \Pr(h_2)\Pr(e'/h_2) + \dots \\ &+ \Pr(h_k)\Pr(e'/h_k)]. \end{aligned} \tag{11}$$

In the statistical case, diverse evidence (as defined by **(B)**) confirms better than does more narrow evidence only if (11) is satisfied.

That inequality (11) is a necessary condition for Horwich to reproduce our intuitions about the superior value of diverse evidence in a statistical context is fatal to his account. Roughly speaking, the left-hand side of (11) describes the effect of conditionalizing on h_1 on the ratio of the probability of diverse to similar evidence. Similarly, the right-hand side of (11) describes a weighted average effect of conditionalizing on h_1 through h_k on the ratio of the probability of diverse to similar evidence. Now **(B)** implies that the right-hand side of (11) is less than one, and it is likely the case that the left-hand side of (11) is also less than one—from **(B)** and the fact that the prior probability of the hypothesis under test is usually significant. Nothing said so far, however, constrains their *relative* ratios. Nor should it, for there seems to be no reason related to the methodological value of diverse evidence for (11) to hold.

It is straightforward to construct cases in which e is more diverse than e' (as defined by **(B)**), but (11) is *not* satisfied. For instance, consider a simple situation in which only three hypotheses have substantial prior probabilities, $\Pr(h_1) = 0.2$, $\Pr(h_2) = 0.2$, and $\Pr(h_3) = 0.6$, and two data sets e and e' such that

$$\begin{array}{ll} \Pr(e/h_1) = 0.2 & \Pr(e'/h_1) = 0.6 \\ \Pr(e/h_2) = 0.4 & \Pr(e'/h_2) = 0.5 \\ \Pr(e/h_3) = 0.4 & \Pr(e'/h_3) = 0.6. \end{array}$$

This is plainly a paradigm case of **(B)**: for all h_j , the $\Pr(e/h_j)$ are significantly less than the $\Pr(e'/h_j)$. Yet a straightforward substitution shows that (11) is violated! Thus we obtain the counterintuitive result that the *similar* evidence lends a greater boost to the hypothesis under test than does the *diverse* evidence. In a uniformly statistical context, then, Horwich's account fails to reproduce our most basic intuition about diverse evidence.

4. Conclusion. To label Bayesian accounts of diverse evidence "success stories" as Earman does is clearly premature. At best, the correlation approach is crucially incomplete, offering a Bayesian account of only one part of our intuition about diversity, while the eliminative approach does not appear to get off the ground. If the considerations of this paper are correct then a sound Bayesian explanation of the superior value of diverse evidence has yet to be advanced, although nothing said so far speaks against the possibility of such an account.

The task for the Bayesian, however, becomes even more formidable if one takes seriously C. Glymour's contention that in large part the methodological value of diverse evidence has nothing to do with correlation or eliminativist intuitions. According to Glymour (1980, 139–142), a body of evidence is diverse to the extent that it enables a hypothesis to be tested in as many ways as possible. Glymour (*ibid.*, 308–309) develops, in the context of his "bootstrap" approach to confirmation theory, formal methods to judge quantitatively diversity of evidence in the context of testing systems of equations.¹ The methodological value of diverse evidence is that it reduces the chance of a spurious agreement between an hypothesis and a body of evidence, thereby reducing the chance that an hypothesis will be confirmed yet be wrong. On this account, variety of evidence has little to do either with the extent to which we judge two pieces of evidence to be correlated, or with the existence of competing plausible hypotheses. Some Bayesians have claimed that all that is right in Glymour's bootstrap theory of confirmation can be integrated directly into Bayesian confirmation theory (Edidin 1983, Rosenkrantz 1983). The problem of diverse evidence calls this claim into question. One of the virtues of the bootstrapping approach is that it gives an account (however preliminary) of one important methodological virtue of variety of evidence. And we have no suggestions yet about how to incorporate this facet of diverse evidence into a Bayesian framework.

The lack of a Bayesian account of diverse evidence stands as an example of a way in which Bayesian methods are less successful than they are advertised to be. By showing Bayesians' claims of success to be over-

¹I am not suggesting that Glymour's approach—or indeed any other formal approach—can render precise and explain variety of evidence better than do the Bayesians. Glymour, for one, is clear that his formal measures of diversity are only preliminary.

stated with respect to the problem of diverse evidence I hope to have made Bayesians' "proof in the pudding" response to global criticisms a little harder to swallow.

REFERENCES

- Earman, J. (1992), *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.
- Edidin, A. (1983), "Bootstrapping Without Bootstraps", in J. Earman, (ed.), *Minnesota Studies in the Philosophy of Science*. Vol. 10, *Testing Scientific Theories*. Minneapolis: University of Minnesota Press, pp. 43–54.
- Franklin, A. and C. Howson, (1984), "Why Do Scientists Prefer to Vary Their Experiments?", *Studies in the History and Philosophy of Science* 15: 51–62.
- Glymour, C. (1980), *Theory and Evidence*. Princeton: Princeton University Press.
- Horwich, P. (1982), *Probability and Evidence*. Cambridge, England: Cambridge University Press.
- Howson, C. and P. Urbach, (1989), *Scientific Reasoning: The Bayesian Approach*. La Salle, IL: Open Court.
- Kyburg, H. (1978), "Subjective Probability: Criticisms, Reflections, and Problems", *Journal of Philosophical Logic* 7: 157–180.
- Rosenkrantz, R. (1983), "Why Glymour is a Bayesian", in J. Earman, (ed.), *Minnesota Studies in the Philosophy of Science*. Vol. 10, *Testing Scientific Theories*. Minneapolis: University of Minnesota Press, pp. 69–97.