



Philosophy of Science Association

Bayes and beyond

Author(s): Geoffrey Hellman

Source: *Philosophy of Science*, Vol. 64, No. 2 (Jun., 1997), pp. 191-221

Published by: The University of Chicago Press on behalf of the Philosophy of Science Association

Stable URL: <http://www.jstor.org/stable/188305>

Accessed: 16/03/2010 16:39

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. 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/action/showPublisher?publisherCode=ucpress>.

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.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



The University of Chicago Press and Philosophy of Science Association are collaborating with JSTOR to digitize, preserve and extend access to *Philosophy of Science*.

Bayes and Beyond*

Geoffrey Hellman†‡

Department of Philosophy, University of Minnesota

Several leading topics outstanding after John Earman's *Bayes or Bust?* are investigated further, with emphasis on the relevance of Bayesian explication in epistemology of science, despite certain limitations. (1) *Dutch Book arguments* are reformulated so that their independence from utility and preference in epistemic contexts is evident. (2) The Bayesian analysis of the *Quine-Duhem problem* is pursued; the phenomenon of a "protective belt" of auxiliary statements around reasonably successful theories is explicated. (3) The Bayesian approach to understanding the superiority of *variety of evidence* is pursued; a recent challenge (by Wayne) is converted into a positive result on behalf of the Bayesian analysis, potentially with far-reaching consequences. (4) The condition for applying the *merger-of-opinion results* and the thesis of *underdetermination of theories* are compared, revealing significant limitations in applicability of the former. (5) Implications concerning "*diachronic Dutch Book*" arguments and "*non-Bayesian shifts*" are drawn, highlighting the incompleteness, but not incorrectness, of Bayesian analysis.

1. Introduction. In his *Bayes or Bust?*, John Earman took great strides in providing a critical evaluation of Bayesianism, as offering "the best good hope for a comprehensive and unified treatment of induction, confirmation, and scientific inference." As he was well aware, on a number of crucial issues the reader would be left hanging: the truth is simply not simple, and we are unsure just how and how far the Bayesian approach can be extended and adapted to illuminate key aspects of scientific method. A short list of leading topics requiring further work includes, among others: (1) the role of Dutch Book arguments and the connection with a relevant notion of "rationality"; (2) the

*Received January 1996.

†Send reprint requests to the author, Department of Philosophy, College of Liberal Arts, University of Minnesota, 355 Ford Hall, 224 Church St. S.E., Minneapolis, MN 55455.

‡I am grateful to Colin Howson, Philip Kitcher, and Paul Teller for helpful correspondence, to Elliot Sober and Alfred Schramm for helpful discussion, and to a referee for critical comments.

Philosophy of Science, 64 (June 1997) pp. 191–221. 0031-8248/97/6402-0001\$2.00
Copyright 1997 by the Philosophy of Science Association. All rights reserved.

Quine-Duhem problem (selective disconfirmation); (3) accounting for the virtues of variety of evidence; (4) the scope and force of the merger-of-opinion results (Doob, Gaifman-Snir); and (5) the status of non-Bayesian shifts of probability functions (reflecting degrees of belief) associated with new theoretical understanding both in normal and non-normal scientific episodes. In what follows, we will try to contribute something further on each of these. In the case of (1), the main point is that Dutch Book arguments can be entirely separated from questions of utility. This is not a new idea, but, I think, it deserves to be emphasized further, especially as it appears to have implications for the last topic (5). Concerning (2), I think the Bayesian analysis can be pushed a bit further to illuminate in a general way the circumstances under which a theory can be protected against disconfirmation *vis-à-vis* auxiliary statements. As to (3), I will in effect reply to a recent critical assessment by Andrew Wayne (1995): although to be sure the Bayesian analysis is incomplete, the situation illustrates one of Earman's main themes, that the Bayesian approach displays a resilience in turning seemingly damning criticisms into valuable insights. On (4), the relationship between the condition of "observational distinguishability" needed for the merger-of-opinion theorems and underdetermination of theories will be re-analyzed; the result is mixed, with some discouraging news along with a glimmer of hope. Finally, on (5), some implications concerning the status of so-called "diachronic Dutch Book arguments" and of "non-Bayesian shifts" will be drawn. What emerges is perhaps a clearer recognition of the essential incompleteness and idealized character of the Bayesian framework but not of any fundamental incorrectness.

2. Dutch Book without Utility: "Dutch Flow" Arguments. If we agree to develop an idealized model of degree of belief by assigning real numbers to propositions or sentences of a given language, we may then appeal to Dutch Book theorems and their converses to argue that those assignments are subject to a Dutch Book if and only if an axiom of the probability calculus is violated. But what does this really establish and what assumptions are really required? It depends on the metatheoretical purpose and context. Three distinct enterprises must be distinguished and rigorously kept separate: (1) the essentially psychological enterprise of assessing actual agents' belief functions, i.e., fixing what those functions are, determining whether they are coherent, etc.; (2) the decision-theoretic enterprise of attempting to ground decisions, courses of action, etc., as "rational" in a certain decision-theoretic sense—call it "rational_d", for explicitness—such that, at a minimum, one is not rational_d if one is prepared, *ceteris paribus*, to be "bilked by a

bookie,” “lose one’s shirt,” etc.; (3) the confirmation-theoretic enterprise of attempting to ground beliefs, methods of testing, acquiring, altering beliefs, etc., as “rational” in an epistemic sense—“rational_e”—further explication of which is one of the goals of the metatheoretic inquiry. I will assume that, whatever their interesting interconnections, these three enterprises are fundamentally distinct and that it is context (3) that is our focus in epistemology of science. To assess a high degree of belief, say, in the atomic hypothesis as rational_e on a given body of evidence by itself implies nothing as to the rationality_a of any course of action whatever, even a course that may affect belief in that very hypothesis. Without further assumptions involving goals, preferences, utilities, or desires, no such inference can be drawn. And to assess such a belief as rational_e does not imply that any betting *behavior* at all is or will be exhibited by any actual agents, even though the (subjective) probability function is understood in terms of fair betting quotients. This latter independence is basic to understanding the role of Dutch Book arguments for context (3), and although it has been explained clearly in the literature—especially by Howson and Urbach (1989)—it is worth repeating and reenforcing the point here.¹

Howson and Urbach (1989) begin their discussion of the theory of betting odds thus:

Our point of departure is the theory of betting odds, and in particular those odds on a hypothesis *h* which, so far as you can tell, would confer no positive advantage or disadvantage to anyone betting on, rather than against, *h* at those odds, on the (possibly counterfactual) assumption that the truth-value of *h* could be unambiguously decided. Such odds, if you can determine them, we shall call your *subjectively fair* odds on *h*. (pp. 56–57)

Soon thereafter they emphasize that judgments of fairness do not imply actual fairness, nor do they imply that anyone will actually make or take bets at subjectively fair odds. “To believe odds to be fair is to make an intellectual judgment, not, except possibly in special conditions which will vary from individual to individual, to possess [even] a disposition to accept particular bets when they are offered” (p. 57). On this basis, it is then claimed that the importance of Dutch Book arguments is this: “betting quotients which do not satisfy the probability

1. What we are labelling rationality_a is similar to what is called “prudential rationality,” e.g., by Pollock (1986), who finds the whole notion of subjective probability incoherent due to a conflation of prudential with epistemic rationality. What we are maintaining here, along with Howson and Urbach, is that there need be no conflation, and that it is possible to read or recast Dutch Book arguments so that it is clear that their object is purely epistemic.

axioms cannot consistently be regarded as determining fair odds.” (p. 59) Thus, an incoherent belief system—one that by definition violates a probability axiom and so is subject to a Dutch Book—*reveals a logico-mathematical contradiction*: the believer has assessed as fair a set of odds that are provably not all fair, by logic and elementary mathematics. This is irrationality_e if anything is, and we need not take up further considerations of utility, preferences, or irrationality in any other sense, such as whether I am irrational_d if I am prepared to lose my shirt or my shoes, etc. (And the converse Dutch Book theorems demonstrate that there is not this kind of inconsistency in a belief system conforming to the probability axioms.)

This freeing of Dutch Book arguments from both betting behavior and utility has welcome consequences, as the well-known vagaries of both are set to one side at a stroke. It does not matter whether I—or my rationally reconstructed counterpart—am actually prepared to enter into any particular bets under given circumstances, whether or not I enjoy or loathe risk—whether my utility of money function is linear—, etc. What matters is that a certain set of judgments—assessments of fairness of certain contractual arrangements—are consistent. The only role of “preference” is one perfectly appropriate to scientific rationality: we prefer not to contradict ourselves! But since “fairness” itself has been framed in terms that seem to involve “values” attached to propositions, there may be lingering doubts as to whether complete freedom from considerations of utility—and from controversial additional assumptions, such as value-additivity with respect to sets of bets—has really been achieved. The following remarks are aimed at dispelling any such lingering doubts.

As usually described, a *bet* on a proposition A , is a contractual arrangement in which a bettor agrees to pay a bookie y monetary units if A turns out to be false and the bookie pays the bettor x units if A turns out to be true. The sum $x + y$ is called the *stakes* of the bet, and the ratio y/x is called the bettor’s *odds*. By definition, the *expected value* or (*expected* or *net*) *advantage* of the bet to the bettor is $x\Pr(A) - y\Pr(\neg A)$, where \Pr is the bettor’s degree-of-belief function. By definition, again, the bet is said to be *fair* just in case this quantity is zero, i.e., just in case there is no expected value or gain on either side of the bet. By the negation principle, that $\Pr(\neg A) = 1 - \Pr(A)$, the condition for a fair bet is that $\Pr(A) = y/(x + y)$; this is called the bettor’s *fair betting quotient*. If the bet is fair, the bettor’s odds y/x are also called *fair*. What the Dutch Book arguments assume, in these terms, then is that (i) the net advantage of a finite (denumerable) set of fair bets is zero (countably infinite sets entering into the arguments for countable additivity); and (ii) that a set of bets, based on given odds, which is

guaranteed to result in a net loss (or gain) in all possible circumstances cannot have zero net advantage, hence not all the odds can be fair.

Looking back, now, at Howson and Urbach's "point of departure," we notice that my assessment that certain odds on a hypothesis h , say, are fair is tantamount to my judgment that *anyone*—i.e., any *hypothetical* bettor—betting *on* or *against* h at those odds would confront zero expected advantage. This says nothing about whether *I* or anyone would actually be *willing* to bet at those odds. (Presumably I would not since my expected gain is exactly balanced by my expected loss, yet I would have to waste energy and time in the transaction. On the other hand, if I were motivated to frustrate such reasoning, perhaps I would bet just for that irrelevant reason.) Indeed, the whole scenario is just a *Gedanken-experiment* designed to help me pinpoint my degree of belief in the proposition in question in numerical terms. Interestingly, I am able to do this—to assess given odds as more or less fair—quite apart from committing myself to any course of action. Even if my risk-aversiveness would discourage me from betting at certain odds, I can judge that I—or anyone else—stand a better-than-even chance to gain at those odds, etc.

Now if this is really the role of betting scenarios, it can be performed by any number of other scenarios which we would not normally describe as involving bets at all. What matters is that I estimate the "expected flow" of some "test quantity" either in the direction of truth of h or of falsity of h . Indeed, there need be no "value" in the ordinary sense attached to the stakes at all. The stakes could be worthless or worse (sand or manure if you prefer, or better an ideal, continuous fluid), so long as it is measured as a real-valued, indefinitely divisible quantity. We can even stop talking about bets entirely and refer instead to "belief tests" ("degree-of-belief tests"). Literally these are just hypothetical set-ups in which a "test fluid" is said to display *positive, negative, or zero expected flow* according as the quantity $x\Pr(h)$ is $>$, $<$, or $= y\Pr(\neg h)$. Here x is the quantity of the fluid that flows "towards h " (or in an arbitrarily chosen "positive direction") if h turns out true, and y is the quantity that flows in the "opposite direction" if h turns out false. $y/y + x$ is now called the "*fair or neutral belief-test quotient*" when the expected flow is zero, i.e., when

$$x\Pr(h) = y\Pr(\neg h). \quad (\text{Zero Expected Flow})$$

And "Dutch Book arguments" can now be called "Dutch Flow arguments": they establish that if and only if a set of belief-test quotients satisfy the axioms of probability can all of them consistently be judged "fair" or "neutral." In these terms, what is required for the Dutch Book construction for the additivity axiom is that the *net expected flow* for

certain *sets* of belief tests “taken simultaneously” (in the finite case, for propositions of the form h_1 , h_2 , and $h_1 \vee h_2$, with h_1 and h_2 mutually exclusive) be computed as the sum of the expected flows for each member of the set. This stipulation of “belief-test-wise additivity” of expected flows corresponds to Howson and Urbach’s assumptions governing “net advantage” of sets of bets.²

There is nothing new here except a relabelling that once and for all strips away all considerations of utility and preference so that the question of consistency is highlighted unadorned. Dutch Flow arguments are Dutch Book arguments in their pure form. This, I submit, is the form appropriate to the context of epistemology.

Behaviorists can be expected to bemoan an alleged loss of contact between assessments of “zero advantage”—which has become “zero expected flow”—and genuine personal conviction. Perhaps honesty and accuracy cannot be assured through thought-experiments involving, say, a noxious fluid, and one should really have to “put one’s money where one’s mouth is” (in the words of Kyburg 1983). Here the disadvantages of a noxious substance will be apparent to just about everyone! But first, honesty and accuracy cannot be guaranteed no matter what medium is chosen. All the problems with behavioral elicitation of degrees of belief arise in any case. Moreover, such problems are a distraction, since our context is (3), not (1), the psychology of belief and belief behavior, something we have set to one side at the outset. Finally, once we start down the path of accurate elicitation, we inevitably become enmeshed in the morass of preference and utility, for these matters infect any behavior that might guide us in assessing personal probabilities. As Earman summed it up (1992, 43–44), “Degrees of belief and utilities have to be elicited in concert.” I would only repeat that utilities constitute another distraction, pertaining as they do to context (2), rationality_a, not epistemology.

This way of looking at matters yields some consequences worth noting. First, the question of “the package principle” is bypassed. Ac-

2. Our stipulations assure, as in standard Dutch-Book arguments, that \Pr is functional, assigning unique values to h (cf. Howson and Urbach, pp. 61–62). Also, to preserve the analogy with standard formulations, the quantities x and y are inverse to $\Pr(h)$ and $\Pr(\neg h)$, respectively. We could, more directly, ask that a fixed quantity, z (the stakes), of indefinitely divisible “test fluid” be divided in the ratio $\Pr(h)/\Pr(\neg h)$ so that this = y/x , i.e. change the “signs” of y and x . Assuming that just whenever zero net flow obtains on the original formulation, this latter identity of ratios also holds, we can already at this stage extract an argument for the negation principle: assigning whatever portion of z that does not flow “toward h ” to “toward $\neg h$ ” simply reflects the *meaning* of classical exclusion negation. This suffices to rule out non-additive “probabilities” assigned to h and $\neg h$, and standard Dutch Book constructions can then be adapted to guarantee additivity of probabilities in general.

ording to this principle, a person's preferences are inconsistent if there is a finite series of bets such that each is regarded as preferable to the status quo while at the same time the status quo is regarded as preferable to the whole series of bets. Schick (1986) has pointed out that this rests on *value additivity*, that the value of a package of bets is the sum of the values of the individual bets. Whatever the merits of this and the package principle in contexts (2) and (1), it is a non-issue in context (3). As we have seen, preference is not involved, and so neither is "value" in any decision-theoretic sense. What remains is merely what we called "belief-test-wise additivity" of expected flows. Thus, there simply is no assumption of value additivity.

A second implication concerns the potential conflict, noted by Earman (1992, 41), between Dutch Book justification and strict conditionalization: the latter leads to cases in which one assigns $\text{Pr}(A) = 1$ to a statement, A —say, a piece of learned evidence—although A is a contingent matter that may fail in some logico-mathematically possible worlds. "As a result," Earman writes (*ibid.*), "something almost as bad as Dutch book befalls the conditionalizer: namely, she is committed to betting on the contingent proposition A at maximal odds, which means that in no possible outcome can she have a positive gain and in some possible outcome she has a loss (a violation of what is called *strict coherence*)." Now from the standpoint of decision theory, violating strict coherence may indeed be "almost as bad as Dutch Book," and this may be a compelling reason for striving to have a "strictly coherent preference structure." But from the standpoint of context (3) and Dutch Flow arguments, the situation is quite different: coherent but not strictly coherent probability assignments involve "almost a contradiction"—that is, really, *not* a contradiction. Since Dutch Book cannot be made, in any sequence of bets there is the possibility of zero net flow, compatible with the assessments of "fairness." The fact that things might in fact turn out badly (or well)—better, flowing to the left or to the right—is irrelevant, as it should be.

There are further implications for diachronic Dutch Book arguments, but it will be better to return to these in the final section below.

3. "Who Deserves the Blame?" On the Quine-Duhem Problem. For the machinery of modern Bayesianism to earn its keep in epistemology of science, it should be able to illuminate scientific method in the sense of providing a rationale for important methodological principles or maxims that have been operative in our best scientific practice and, if possible, leading us to sharper formulations of such maxims and even to new ones of proven or potential scientific value. Perhaps a paradigm case is the explication of *severity of tests*, which we can quickly review

as an example of the sort of achievement that is both useful and within Bayesian means.

Let evidence statements describing the passing of tests of a hypothesis H be labelled E_i , all relative to a fixed background of accepted knowledge K . Then we *define*:

E_1 is a more severe test of H rel. K than $E_2 \equiv \Pr(E_1/\neg H \ \& \ K) < \Pr(E_2/\neg H \ \& \ K)$

that is, the first test is less likely to be passed given the falsity of the hypothesis than the second is. Note that this is a comparative notion; we are not defining severity in any absolute sense.³ Given this definition, we have the following

THEOREM: If the likelihoods of E_1 and E_2 given $H \ \& \ K$ are comparable, then: E_1 results in greater incremental confirmation of H on K than E_2 if and only if E_1 is more severe than E_2 .

PROOF (Earman style): Write out Bayes' Theorem. (Recall that incremental confirmation of H on K by E is defined as the difference $\Pr(H/E \ \& \ K) - \Pr(H/K)$.)⁴

Do not complain that the proof is a mathematical triviality and that our definition of greater severity provides us with no method for determining in general when one test is more severe than another. An easy proof is an advantage, so long as it is not based on an epistemic circularity. And it is usually too much to require that a definition provide us with a method of deciding cases, especially in a context as broad and "non-computational" as epistemology of science. It is an advance if a definition—or perhaps only a partial definition—can be given which (i) fits with a sizable body of particular presystematic judgments, positive and/or negative, depending on the use to which the definition is put, and (ii) enables us to forge a link with other important concepts which have been independently explicated (in this case, "degree of confirmation"). Moreover, if the definition makes us aware of an epistemological problem—e.g., assessing the likelihood, $\Pr(E/\neg H)$, where $\neg H$

3. We are thus side-stepping problems that Earman raises (1992, 116–117) concerning an absolute criterion of severity, as appealed to by Mayo (1991).

4. Without claiming that this choice of measure of degree of confirmation is optimal, we would suggest that it is an intuitive one and superior in this regard to the ratio measure— $\Pr(H/K \ \& \ E)/\Pr(H/K)$ —which is sometimes proposed. On the latter, evidence which raises the credibility of a hypothesis from "one in a million" to "one in a thousand" is "more confirmatory"—indeed "(more than) ten times more confirmatory"—than evidence that raises credibility from "one in a hundred" to "ninety-nine out of a hundred." Similar considerations apply to measures of "probabilistic relevance" of hypotheses to data, cf. below, Section 4.

may be broken down into a disjunction, one of whose disjuncts remains uncharted territory for us—that may be a virtue, not a defect. And, of course, it is hardly a strike against an explication that it calls for further analysis, refinement, and application to special classes of cases.

Earman describes a qualified success of this sort regarding the Quine-Duhem problem (1992, 84–85). Here is another. It helps flesh out the Bayesian claim to explicate the phenomenon of a “protective belt” of auxiliary statements around entrenched theories, highlighted by Lakatos. (Cf. Howson and Urbach 1989, 94–102.)

To help fix ideas, let N stand for Newtonian mechanics used in predicting planetary orbits, let E stand for predictions of Uranus’ orbit violated by observations \tilde{E} which led to the Adams-Le Verrier discovery of Neptune,⁵ and let A stand for the auxiliaries used to obtain E from N , including crucially the “extremal condition” that “no other substantial masses are in the vicinity of Uranus” (call this A^*). Finally let K abbreviate background knowledge, including all the accumulated evidence in support of N (available to the communities of Adams and Le Verrier, say). It is widely regarded as reasonable to have blamed A , especially A^* , more than N for the failure of E , \tilde{E} , even prior to the discovery of Neptune. The search for Neptune was a highly reasonable course, not a dogmatic effort to save a theory. Confidence in its existence was considerable; the difficulty was largely mathematical, to predict its location accurately so that it could be found in a search with available telescopes. What can the Bayesian say, in a general way, about such situations?

We have $N \& A \vdash E$ and $\tilde{E} \vdash \neg E$, which will be assumed throughout, whence

$$(1) \quad \Pr(A/N \& \tilde{E}) = 0 \text{ (as well as } \Pr(N/A \& \tilde{E}) = 0).$$

Assume also that $\Pr(K) = 1$, in particular,

$$(2) \quad \Pr(K/N) = 1.$$

Now define

$$(3) \quad \varepsilon_N = \Pr(K) - \Pr(N \& K) (= 1 - \Pr(N/K)).$$

The following lemma is readily proved from elementary probabilistic algebra:

LEMMA: (1)–(3) $\Rightarrow \Pr(A \& \tilde{E} \& K) \leq \varepsilon_N$.
(Proof in Appendix.)

5. For a bit of the fascinating history of Adams’ role, see Fernie 1995.

Now, intuitively, we are tempted to say that, because the antecedent support of N by K was large— K “almost implies” N —and since N & \tilde{E} preclude A , therefore K & \tilde{E} “almost preclude” A . (Note that, in the historical example, some of A may overlap with K , but the dubious part, A^* , certainly was unknown, so the parallel inference interchanging ‘ N ’ and ‘ A ’ is not tempting— K doesn’t “almost imply” A .) Notoriously, this kind of reasoning is not generally correct in probabilistic logic—the difference between measure 0 and positive measure can thoroughly upset such inferences. But the above lemma can still be used to set some limits.

Rather than look at the incremental (dis)confirmation of N and of A by \tilde{E} , we may concentrate on just the comparison of their posteriors. Let us define a theory T to be *protected against counterevidence* \tilde{E} vis-à-vis auxiliaries A (rel. K)—where $T, A \vdash E$ and $\tilde{E} \vdash \neg E$ —to mean

$$\Pr(T/K \ \& \ \tilde{E}) > \Pr(A/K \ \& \ \tilde{E}).$$

This allows that both T and A may suffer by various amounts in the face of \tilde{E} , but protection occurs if T ends up better off than A —even if it suffered more in the process.

Concerning notation: As an aid to seeing the significance of formulas, let us continue to use ‘ N ’, ‘ \tilde{E} ’, etc. as general variables, except when explicit reference to the historical example is indicated.

Now notice that an equivalent to protection of N against \tilde{E} vis-à-vis A rel. K is:

$$\Pr(\tilde{E}/K) > \frac{\Pr(A \ \& \ K \ \& \ \tilde{E})}{\Pr(N/K \ \& \ \tilde{E})}. \quad (\text{Pro})$$

(Interchange the left side with the denominator of the right and observe that, since K is assumed known,

$$\Pr(A/K \ \& \ \tilde{E}) = \Pr(A \ \& \ K \ \& \ \tilde{E})/\Pr(\tilde{E}/K).)$$

Now an immediate corollary of the Lemma is

COROLLARY: A sufficient condition for protection (Pro) is that

$$\Pr(\tilde{E}/K) > \frac{\varepsilon_N}{\Pr(N/K \ \& \ \tilde{E})} = \frac{1 - \Pr(N/K)}{\Pr(N/K \ \& \ \tilde{E})}.$$

This can be put as

Maxim 1: The more highly N is confirmed “prior to” \tilde{E} , and the less (negative) impact \tilde{E} has on N in the absence of A less K , the smaller must $\Pr(\tilde{E}/K)$ be for protection to fail.

In the paradigm example, not only was N more highly confirmed apart

from the anomalous \tilde{E} than A^* , but there were also various mechanisms—within and without the framework of N —that could have produced \tilde{E} , so that $\Pr(\tilde{E}/K)$ was not negligible. The remaining factor, the posterior $\Pr(N/K \ \& \ \tilde{E})$, was also arguably not significantly different from the prior $\Pr(N/K)$, as we shall see momentarily.

Maxim 1 depends only on our stated assumptions and definitions and not yet on Bayes' Theorem. Of course, Bayes' Theorem can be expected to enter in estimating the posterior $\Pr(N/K \ \& \ \tilde{E})$ occurring in the denominator in the display of the Corollary. A special case of this is also inspired by the paradigm example. Notice that, without any uncertain auxiliary assumptions, Newton's laws were silent on the question of \tilde{E} : N gave no further information about \tilde{E} not already available in K . Thus, we may reasonably write $\Pr(\tilde{E}/K \ \& \ N) = \Pr(\tilde{E}/K)$. Now Bayes' Theorem tells us immediately that

$$\Pr(N/\tilde{E} \ \& \ K) = \Pr(N/K).$$

The sufficient condition for protection in the above Corollary then takes the form, $\Pr(\tilde{E}/K) > \epsilon_N/\Pr(N/K) = \epsilon_N/(1 - \epsilon_N)$. But we can do better without even invoking the Lemma. For if $\Pr(N/K) > .5$, then under the conditions of the Lemma, it readily follows that $\Pr(A/K \ \& \ \tilde{E}) \leq 1 - \Pr(N/K \ \& \ \tilde{E}) < .5$, in the special case, so that N is protected. This helps explicate and support the following, as a corollary of Bayes' Theorem:

Maxim 2: If a theory N is more likely than not (on background K) and gives no information as to \tilde{E} in the absence of uncertain auxiliaries A , then it is protected against \tilde{E} vis-à-vis A ; indeed $\Pr(N/K \ \& \ \tilde{E}) = \Pr(N/K) > .5 > \Pr(A/K \ \& \ \tilde{E})$.

These maxims state sufficient, not necessary, conditions for protection in the sense defined. The probabilistic relations that underwrite them provide some measure of the wide range of circumstances in which a reasonably successful theory can be protected against "anomalous" observations, of how hard it may be to reduce confidence in such a theory to or below that of uncertain auxiliaries, much less "overthrow" the theory in a Kuhnian sense. As in the case of Earman's analysis, objectification of these results is by no means automatic, and depends in any given case on the ways in which the priors involved have been reached. Still, they go some distance—beyond the mere description of examples—toward providing an epistemological rationale for aspects of scientific practice that too readily have been seen as manifestations of dogmatism or of purely non-rational social forces. And, in any case, they have a certain normative force: to the extent that the

relevant priors have been reasonably assigned, protection of a theory against “refutation” under the stated circumstances is also reasonable.

In order for analyses such as the above to be developed into a true Bayesian “success story,” it will be necessary to say much more about the hard cases of theory choice in revolutionary circumstances, in which an entrenched old theory, fairly successful in its domain, confronts anomalies and a rival new theory which may overcome some of these but which is itself still in a rather nascent state (e.g., early Copernicanism, early Darwinianism, etc.). This is broader than the standard Quine-Duhem problem context, since one is comparing two or more bodies of theory with their own sources of support, with their own anomalies or other difficulties, and perhaps with distinctive standards and criteria of success. A full-scale Bayesian treatment will probably require putting together several pieces corresponding to various aspects of good method. (A valiant attempt to display the superiority of Copernican over Ptolemaic astronomy in a single, telling Bayesian mechanism has been found problematic. See Gardner’s (1983) criticisms of Rosenkrantz 1977.) Although such a synthesis is well beyond the scope of this paper, one of the most important aspects of good method, widely cited on behalf of succeeding revolutionary new sciences, can be addressed in Bayesian terms. Or so we shall now argue.

4. Variety of Evidence. One of the most significant virtues a good scientific theory can possess is a capacity to account for a wide variety of phenomena hitherto regarded as disparate. Wide explanatory scope is clearly a component of unifying power, and unifying power in turn is closely related to explanatory success.⁶ The intuitive idea that diverse or more various evidence has greater confirmatory force than routine or ordinary evidence—other things being equal—is thus important not only in its own right but also in helping forge a potentially interesting link between explanation and confirmation. To the extent that the forging of this link is successful, a kind of rationale *cum* explication of “inference to the best explanation” (“IBE”) is also in the offing; placing greater confidence in a theory which explains better often reduces to a recognition of greater posterior probability due to the effects of confirmation by more various evidence. In this sense, at least some forms of IBE do not go outside the general framework of probabilistic inference.⁷

6. Indeed, on the comprehensive approach to explanation developed by Kitcher (1989), inspired by Friedman (1974), unification is the essence of explanatory success.

7. This stands in contrast with certain ways of understanding “inference to the best explanation” criticized by van Fraassen (1989, Ch. 6 and 7).

Can Bayesian efforts to underwrite the virtues of variety of evidence be developed into a “success story”—potentially a “big success story,” if the above reflections are not wide of the mark? Earman provides some grounds for optimism, reviewing two known “mechanisms” whereby more various evidence confirms more than a comparable amount of ordinary or “similar” evidence. The first of these, the “correlational approach,” rests on the assumption that more various pieces of evidence are less well correlated with one another than more similar pieces, and this shows up in a slower growth of the denominator of Bayes’ formula for the posterior of the hypothesis on the whole body of evidence, i.e., a greater boost in the posterior on the more various as compared with the more similar evidence. The second mechanism, “efficient elimination of rival hypotheses,” also appeals to Bayes’ formula, written in terms of a partition of the theory (hypothesis) space, and rests on the assumption that more various evidence will tend to reduce the probability of rival hypotheses, as many of them will provide lower likelihoods to such a body of evidence than to a body of similar evidence.

Recently, both these mechanisms have come under criticism by Andrew Wayne (1995). On analysis, it is granted that the first mechanism can be redescribed so as to sustain the

important consequence that diverse evidence confirms better than does the same amount of similar evidence when the individual priors are comparable (Wayne 1995, 115),

but the account is found “incomplete” for failing to account for the dependence of judgments of diversity on changing theoretical context and background knowledge. The second mechanism is seen as potentially compensating for this incompleteness, and it is conceded that, if it were successful, “it would constitute a large step toward counting diverse evidence among Bayesian success stories.” However, it is argued not to be successful: in the deterministic case (or hypothetico-deductive case of theories assigning 1 or 0 likelihoods to the evidence), the account is found viciously circular. In the non-deterministic, or statistical case, the verdict is even worse: the account is alleged to be fatally flawed (Wayne 1995, 118–119).

I believe these criticisms, while containing a kernel of truth, are ultimately mistaken. In the deterministic case, there really is no problem of definitional circularity, appearances to the contrary notwithstanding. And in the statistical case, as I will try to show, we have a beautiful illustration of Earman’s point concerning Bayesianism’s resilience: the criticism provides an insight which can be turned into a theorem, one that highlights a connection between confirmation and explanation.

In order to obtain any definite Bayesian results, it is necessary to connect informal concepts with formal, probabilistic conditions, as we saw above in the case of “severity of tests.” The connection may take the form of a definition or, for certain purposes, it may involve something weaker, e.g., a necessary or a sufficient condition. In any case, the construction must be judged according to how well it captures important instances of ordinary or intuitive usage, and also according to its fruitfulness as part of the larger enterprise of explication. In the case of variety of evidence, the quasi-formal criterion discussed critically by Wayne is due to Horwich (1982). Suppose we are given a hypothesis h_1 under test and a set of exhaustive and mutually exclusive hypotheses $\{h_1, h_2, \dots, h_k\}$, and 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,

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

whereas there are few if any cases of the converse situation (with the opposite inequality).

In this definition and elsewhere, reference to given background knowledge has been suppressed, presumably for ease of notation. But of course, judgments of variety are dependent on the background, so this relativity is present throughout. Note also that (B) is stated as a necessary and sufficient condition, but in fact what is needed in the use to which it is put is merely that the condition for greater diversity be necessary, that is, given that e counts as more diverse than e' relative to a given background—*however that is ultimately analyzed*—the likelihoods are expected to obey the stated condition. And when this is so, it can be shown that, under a variety of conditions, the more diverse data set e confirms the hypothesis (h_1) more than does the less diverse e' , in the sense that the ratio of posteriors $\Pr(h_1|e)/\Pr(h_1|e') > 1$. (Note that here and below we may stick with the difference measure of degree of confirmation.) This already is enough to rebut the charge of logical circularity: (B) need not be read as a *definition* of greater variety, but merely as a quasi-formal consequence of greater variety. It then becomes an empirical question whether a substantial body of actual data sets of interest—*independently assessed as more or less diverse, perhaps only by intuitive standards*—and hypothesis sets do indeed fulfill the quasi-formal condition.

To see how the arguments work, look first at the case considered by Horwich, in which h_1 is assumed deterministic and the other h_j may be statistical. Expand

$$\Pr(e) = \Pr(h_1)\Pr(e|h_1) + \Pr(h_2)\Pr(e|h_2) + \dots + \Pr(h_k)\Pr(e|h_k),$$

and then substitute into Bayes' Theorem to obtain,

$$\frac{\Pr(h_1|e)/\Pr(h_1|e')}{[\Pr(h_1) + \Pr(h_2)\Pr(e'|h_2) + \dots + \Pr(h_k)\Pr(e'|h_k)] \div [\Pr(h_1) + \Pr(h_2)\Pr(e|h_2) + \dots + \Pr(h_k)\Pr(e|h_k)]} = \tag{7}$$

(following Wayne's numbering). Notice that the right hand side = $\Pr(e')/\Pr(e)$. Wayne then interprets (B) as implying that this ratio $\Pr(e')/\Pr(e) > 1$, and so therefore is the left of (7), which is the desired result.⁸ In the case in which all the hypotheses are deterministic, (B) is replaced by the condition that many of the alternative hypotheses h_j with substantial prior probabilities give $\Pr(e|h_j) = 0$ and $\Pr(e'|h_j) = 1$, with the converse situation a rarity. Then (7) reduces to

$$\Pr(h_1|e)/\Pr(h_1|e') = \Sigma_i \Pr(h_i)/\Sigma_j \Pr(h_j), \tag{9}$$

where i ranges over the hypotheses not eliminated by the similar evidence—i.e., $\Pr(e'|h_i) \neq 0$ —and j ranges over those not eliminated by the diverse evidence. Again the right side is > 1 , giving the desired result.

At this point, Wayne writes:

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. (1995, 118)

But the charge is not just. Diverse evidence *confirms* better because it eliminates more rival hypotheses—that is the content of the *theorem*, that the ratios of (9) are > 1 , based on the “definition” of greater diversity. Two points are in order. First, the theorem is not empty: the mechanics of Bayes' Theorem tell us that “elimination of rivals” in the sense of the condition on *likelihoods*—not directly on posteriors—does indeed result in increased confirmation by the more diverse evidence. Second, as already explained, that is not the end of the matter, for the “definition” required is really only a one-way connection between in-

8. Strictly, (B) implies $\Pr(e')/\Pr(e) > 1$ only if $\Sigma_i \Pr(h_i)[\Pr(e|h_i) - \Pr(e'|h_i)] < \Sigma_j \Pr(h_j)[\Pr(e'|h_j) - \Pr(e|h_j)]$, where i ranges over the (exceptional) h such that

$$\Pr(e|h) > \Pr(e'|h)$$

and where j ranges over the (more common) h such that

$$\Pr(e'|h) > \Pr(e|h).$$

formal usage and a quasi-formal condition, and, moreover, this connection depends on the substantive matter of whether or not a decent match is achieved between informal and quasi-formal usage. The explanation of the virtue of variety is thus perfectly in line with the standard pattern of Bayesian explication, as illustrated above for “severity of tests”: a formal or quasi-formal condition is given as a stand-in for an informal condition, and then the machinery of probability theory—usually involving a form of Bayes’ Theorem—is used to derive certain desired or instructive confirmational consequences.

To be sure, nothing like a “complete analysis” of variety of evidence is achieved. And one would certainly like an account of how judgments of variety depend on theoretical context and background. But this goes beyond the scope of providing a Bayesian rationale for the *superiority* of variety. Moreover, I suggest that the question about changing standards of diversity really concerns the more general question of “objectivity” of the probabilistic judgments. If so, it is appropriate to reply, “One thing at a time!”

Turning to the case of a set of statistical hypotheses, $\Pr(e/h_i) < 1$, condition (7) becomes,

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

Note that the right hand side =

$$\Pr(e/h_1)\Pr(e')/\Pr(e'/h_1)\Pr(e). \quad (10')$$

The condition that the more diverse e confirm h_1 better than the less diverse e' in the sense at issue is that the ratio of the left, hence right, of (10), hence that of (10'), be > 1 . Now, Wayne observes (p. 119), for this to be the case, it is necessary and sufficient that

$$\Pr(e/h_1)/\Pr(e'/h_1) > \Pr(e)/\Pr(e'), \quad (11')$$

which is obtained from Wayne’s formula (11) by substituting $\Pr(e)$ for the bracketed portion of the denominator of (10) and $\Pr(e')$ for the bracketed portion of the numerator of (10). At this point, Wayne writes,

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. [Although (B) implies that the right side of (11), or (11') is < 1 , and it is likely that the left side is too . . .] 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. (1995, 119)

Wayne then presents a contrived example in which e is more diverse than e' (by (B)) but in which (11) is not satisfied, and the more similar evidence e' boosts the hypothesis slightly, whereas the diverse evidence e actually *disconfirms* the hypothesis. But this example has the curious feature that $\Pr(e/h_1) = .2$ whereas $\Pr(e) = .36$, while $\Pr(e'/h_1) = .6$ and $\Pr(e') = .58$. In other words, the hypothesis in question is probabilistically *negatively relevant to*—or we can say (probabilistically) *counter-explains*—the diverse data e , whereas it is *positively relevant to*—or (probabilistically) *explains*—the similar data e' slightly. To be sure, there is nothing to be wrung out of the concept of “variety” all by itself to prevent this and the consequent violation of (11) ((11')), but that is of little interest. After all, we know in advance, by a straightforward application of Bayes' Theorem, that *any* evidence—diverse or otherwise—whose likelihood is lowered by a hypothesis (with non-zero prior), relative to a given background, disconfirms that hypothesis. So if we are interested in the confirmatory virtues of diverse data, we had better focus on cases in which there is *a priori* at least the possibility of confirmation, i.e., cases in which the data counts as evidence, not counterevidence! Moreover, we are anyway principally concerned with the virtues of variety of evidence for theories which help explain the data, and this normally requires positive statistical relevance.⁹ What is interesting, then, about condition (11') is that it can be used to generate results such as the following:

If a hypothesis h probabilistically explains a diverse data set e at least as well as it probabilistically explains a similar data set e' , then the diverse data confirms the hypothesis better than does the similar data, in the sense that the ratio of (10) is > 1 .

What is meant here by “explains at least as well” is just that

$$\Pr(e'/h) - \Pr(e') \leq \Pr(e/h) - \Pr(e),$$

where we restrict ourselves to cases in which both sides are non-negative. Note that it is not required that the likelihood $\Pr(e/h)$ be $\geq \Pr(e'/h)$, but merely that the increment given by h to $\Pr(e)$ be as great

9. Cases of explanatory hypotheses which are not positively statistically relevant are well known from discussions of “Simpson’s paradox,” highlighted by Cartwright (1979). Typically, however, positive relevance is restored by suitably disaggregating the statistics (i.e., by conditionalizing on further factors known to be relevant). Cf. van Fraassen 1980, 148–149.

as that given to $\Pr(e')$. The result claimed follows in the simplest sub-case of (B), implicitly considered by Wayne in connection with (7) above, in which

$$\Pr(e)/\Pr(e') < 1, \text{ i.e. } \Pr(e) < \Pr(e'),$$

together with the following elementary fact about real numbers p, q, r , all > 0 :

$$p < q \leftrightarrow (p + r)/p > (q + r)/q.$$

Let p be $\Pr(e)$ and q be $\Pr(e')$ and suppose, for simplicity, that h boosts both by $\varepsilon > 0$, i.e. $\Pr(e/h) = \Pr(e) + \varepsilon$ and $\Pr(e'/h) = \Pr(e') + \varepsilon$. Then

$$\Pr(e/h)/\Pr(e) > \Pr(e'/h)/\Pr(e'),$$

and the moreso as the difference between $\Pr(e')$ and $\Pr(e)$ is greater. But this is equivalent to (11'),

$$\Pr(e/h)/\Pr(e'/h) > \Pr(e)/\Pr(e'),$$

which insures that the ratio of (10) is > 1 , which is the desired result. Obviously the result holds *a fortiori* if the explanatory boost given to e is greater than that given to e' (the latter of which *a fortiori* may be negative). If we call such an incremental boost, $\Pr(e/h) - \Pr(e)$, the “(positive) explanatory force of h for e [relative to the background, as throughout]” then we may state the result as follows:

Corollary of Bayes' Theorem: Let e be more diverse than e' such that $\Pr(e) < \Pr(e')$, and let h be such that its (positive) explanatory force for e is as great as or greater than its explanatory force for e' ; then $\Pr(h/e)/\Pr(h/e') > 1$.

A kind of dual to this is obtained by substituting $-r$ for r and setting $\varepsilon < 0$ in the above and reversing relevant inequalities: If the explanatory force for e is as negative as that for e' or more negative, then the ratio $\Pr(h/e)/\Pr(h/e') < 1$. This covers the situation Wayne encountered; but it goes hand in hand with the positive result, which is surely a victory for the Bayesian approach.

Note furthermore that the positive result is a natural generalization of what happens in the deterministic case. There we had both $\Pr(e)/\Pr(e') < 1$ and $\Pr(e/h_1) = \Pr(e'/h_1) = 1$, which means that the (positive) explanatory force of h_1 for e , the more diverse set, is greater than that for e' , the less diverse one.

An obvious limitation of results such as these is that different bodies of evidence of varying diversity are being compared for their affect on

one and the same hypothesis. But, as was noted above in connection with the Quine-Duhem problem, some of the most challenging questions for Bayesianism involve comparisons among *different* theories (or hypotheses), as in the famous revolutionary episodes. Moreover, a frequently cited factor favoring revolutionary shifts in the famous successful cases (Copernicanism, Darwinism, etc.) is the greater variety of phenomena “accounted for” by the new science.¹⁰ The Bayesian is thus called upon to underwrite the superior force of more diverse evidence e supporting some h_{new} as compared with less diverse e' supporting a radically different h_{old} . A cursory glance at the above analysis leading to the last Corollary reveals that if priors $\Pr(h_{\text{old}}/k)$ and $\Pr(h_{\text{new}}/k)$ are equal, and if the positive relevance of h_{new} for e is as great as or greater than that of h_{old} for e' , then the reasoning leading to the Corollary can be applied to conclude that $\Pr(h_{\text{new}}/k \ \& \ e) > \Pr(h_{\text{old}}/k \ \& \ e')$. But this is not enough, since we must consider the hypotheses on the total evidence, which includes $e \ \& \ e'$. In the starkest case, h_{old} fails utterly to explain the diverse e , whereas h_{new} succeeds as well as h_{old} in explaining the narrow e' . It then follows that $\Pr(h_{\text{new}}/k \ \& \ e \ \& \ e') > \Pr(h_{\text{old}}/k \ \& \ e \ \& \ e')$. One may well ask how *diversity* matters here, since a larger body of (boosted) evidence, diverse or routine, can be expected to boost credibility more than a smaller part. The answer the Bayesian can give is that it is, of course, a matter of degree, and the more diverse the excess of evidence, e , supporting h_{new} beyond that supporting both h_{old} and h_{new} , e' , the better the gain for h_{new} . Indeed, even if the prior $\Pr(h_{\text{new}}/k)$ is less than $\Pr(h_{\text{old}}/k)$, the effect of successive bodies of diverse evidence can drive the probability of h_{new} well past that of h_{old} . The more diverse those bodies are themselves, internally and compared with each other, the more efficiently this occurs, *ceteris paribus*. Thus, the Bayesian need by no means assume equality of the priors, and there would seem to be much that Bayesian analysis can say in favor of the well-known revolutionary sciences on the score of variety of evidence alone. Here as elsewhere, there is the twin problem of identification and objectification of the priors in any given case, and I have no general prescription to offer, other than to suggest that the continuing effort to deploy the Bayesian conceptual apparatus in careful case studies should yield some rewarding results. As to the general question of objectification, that is our next topic.

5. Merger of Opinion and Underdetermination of Theories. To what

10. See, for example, Michael Gardner's (1983) discussion of the Copernican revolution and the importance of variety of evidence in bolstering Copernican arguments. What Bayesianism has to say concerning variety of evidence, however, is not considered.

extent can subjective degree-of-belief functions be “objectified” at least in the sense that differences among them concerning scientific hypotheses at a given stage of inquiry can be made smaller and smaller at later stages with accumulating evidence? Among the most striking results on behalf of Bayesianism are the so-called “convergence to certainty” and “merger of opinion” results, which show essentially that repeated conditionalization on accumulating evidence tends to drive differing Pr functions—equally dogmatic in assigning 0 to the same hypotheses—toward agreement. Some of the leading precise formulations and theorems are reviewed by Earman (in Ch. 6), and I will not reduplicate that review here. I want to call attention to two points, however, affecting the significance of the strong result Earman focuses on, the Gaifman-Snir Theorem, based on martingale convergence theorems. The first point concerns the relationship between underdetermination of theories by “observation” and the condition of “weak observational distinguishability” (“*wod*”) of competing theories needed in order to apply the Gaifman-Snir Theorem to theoretical hypotheses. The point is that, properly analyzed, underdetermination is not equivalent to failure of weak observational distinguishability: *wod* can readily fail without implying underdetermination; however, underdetermination together with a mild model-existence assumption does imply failure of *wod*, i.e., *wod* does rule out underdetermination. The upshot is to reinforce Earman’s qualms about the strength of the distinguishability required for Gaifman-Snir: it is even stronger than his analysis suggests. The second point is a compensating one, concerning the shifting and *expanding* character of “observation” and “evidence” with scientific progress.

For our purposes here, we need not present a precise statement of the Gaifman-Snir Theorem. What concerns us is the formal condition of a set Φ of sentences in a first-order language \mathcal{L} *separating* a subset K of the (mathematically standard) models or structures for \mathcal{L} , $\text{Mod}_{\mathcal{L}}$. This is said to obtain just in case, for any pair of models w_1 and w_2 of K , there is a $\varphi \in \Phi$ such that $w_1 \models \varphi$ and $w_2 \models \neg\varphi$. If Φ is identified as an “evidence matrix” and separates $\text{Mod}_{\mathcal{L}}$ then the Gaifman-Snir Theorem says that, for any sentence ψ of \mathcal{L} , as evidence from Φ accumulates in a world w , the posterior probability of ψ on that evidence approaches 1 or 0, respectively, as ψ is true or false in w , for almost every w (i.e., a Pr 1 set). Moreover, the “distance” between any two equally dogmatic Pr functions in the “uniform distance metric” tends to 0 as the evidence accumulates (i.e., the maximum disagreement between the functions over sentences ψ given the accumulating evidence shrinks to 0).

The role of the separation condition is crucial in allowing applica-

tion of the Doob martingale construction to the probability space based on $\text{Mod}_{\mathcal{L}}$. The theorems as such, of course, say nothing about “evidence,” “observation,” “theory,” etc., but applications to scientific reasoning naturally bring in such notions. If we think of conditionalizing on “evidence statements,” then this corpus must separate the models of our scientific language. If this language is artificially restricted to the vocabulary of the “evidence” corpus itself—if we consider only “empirical hypotheses,” for example—and if the corpus includes all atomic “observation sentences,” then the Gaifman-Snir theorem can be applied to infer convergence to certainty on the truth of the hypotheses in almost every “empirically possible world” (i.e., logical model or structure for the empirical language). This is by no means a trivial result, especially in that the hypotheses in the role of ψ of the theorem may be of arbitrary quantificational complexity. However, as Earman points out, as soon as new predicates, say “theoretical predicates,” are added to the language \mathcal{L} , the original “evidence matrix” will not be separating and the theorem will not apply. To restore applicability, a further condition of “distinguishability of theoretical hypotheses by evidence” must be met.

Following Earman (1992, 151), let T_1 and T_2 be incompatible theories; they are called “*weakly observationally distinguishable*” (*wod*) for the models \mathcal{M} just in case for any w_1 and w_2 in \mathcal{M} such that $w_1 \models T_1$ and $w_2 \models T_2$ there is a (possibly quantified) observation sentence E such that $w_1 \models E$ and $w_2 \models \neg E$. If we consider a partition $\{T_i\}$ of theories such that the theories are pairwise wod for $\mathcal{M} = \text{Mod}_{\mathcal{L}}$ where \mathcal{L} is the union of the languages of the theories, then the evidence statements will serve to separate the models \mathcal{M} and the convergence results will apply to the sentences of the theories: in almost every model, accumulating evidence will drive the probability of such a sentence to 1 or 0 according as it is true or false in the model.

The first thing to notice about the condition wod is that it is equivalent to the following “determination condition” on the models in \mathcal{M} :

If w_1 and w_2 agree on all “empirical sentences” E , then they agree on all theoretical sentences T such that $T \in T_i$, some i . (E-Determ)

In other words, if any two models differ in truth-value assigned to any sentence in theoretical vocabulary which is a theorem of one of the theories, then those models must also differ in some empirical respect. This is thus of the same form as a type of determination relation used to explicate “physicalism” (cf. Hellman and Thompson 1975), also called strong or global “supervenience” (Kim 1984). Intuitively, then, this is quite a strong condition. (It is called “weak” by Earman only to distinguish it from a condition which is even stronger, in which the

empirical sentence(s) corresponding to any theoretical disagreement can be chosen independently of the worlds.) Another useful formulation—nearly enough equivalent—is this: let ‘ w^e ’ denote the reduct of model w to the evidence-vocabulary (think of such reducts as “possible empirical worlds”), and call w an *expansion* (to non-evidence-vocabulary) of w^e ; also, use ‘model of T ’ to mean structure for \mathcal{L} which makes true all theorems of T . Then E-determination says,

Any w^e can be expanded to a model w of at most one of the theories
 T_i , (\approx E-Determ)

(The requirement that $\{T_i\}$ be a partition then implies “exactly one of the theories T_i ”.) That is, no pair of models of different theories have the same reduct to the E vocabulary. In this sense, empirical worlds are “compatible with” at most one of the theories of a partition.

What is the relationship between E-determination of the theoretical in this sense and underdetermination of theory by evidence? Underdetermination in any philosophically interesting sense requires, not merely that there be *some* possible evidential situation or world which fails to discriminate between disagreeing theories, but that *every possible evidential situation* or world so fail to discriminate. This motivates the following definition:

T_1 and T_2 are *E-indiscriminable* iff any E-structure w which can be expanded to a model w_1 of T_1 can also be expanded to a model w_2 of T_2 , and conversely (so that $w = w_1^e = w_2^e$, in the notation for reducts used above). (E-Indiscrim)

This seems to yield a reasonable standard of underdetermination of theory by (possible) evidence: underdetermination obtains just when there are E-indiscriminable incompatible theories in this sense. It fails just in case every incompatible pair of theories is *E-discriminable*, where this is just the negated condition:

T_1 and T_2 are *E-discriminable* iff there exists an E-structure w which can be expanded to a model of T_1 but not to one of T_2 or vice versa. (E-Discrim)

Thus, clearly we have the following relationship:

E-discriminability of theories is *insufficient* for their weak-observational distinguishability.

It can easily be the case that some possible empirical world discriminates between a pair of theories while another fails to, e.g., for want of occurrence *in that world* of any of the experiments that might be needed to do the job. Then the theories are not wod, as every empirical

world (expandable to a model of either theory) must in effect discriminate, by (\approx E-Determ). In other words, we have that

Failure of theories to be wod is *insufficient* for an instance of underdetermination of theories by evidence.

The difference between the two conditions is vast: Failure of wod says merely that *some* possible empirical world fails to discriminate between the theories in question, whereas underdetermination requires that *every* such world fail to discriminate. To take failure of wod as a standard for underdetermination is to allow the latter to obtain *far* too easily.

Suppose conversely, however, that there are instances of genuine underdetermination in the sense of E-indiscriminability. Say the theories involved are just two, labelled T_1 and T_2 . Assume also that there exists an E-structure w which is compatible with—i.e., can be expanded to a model of—at least one of these theories. (This is hardly any requirement at all, as we can assume that all models have empirical substructures.) E-indiscriminability implies that w can be expanded to a model w_1 of T_1 iff it can be expanded to a model w_2 of T_2 . By the assumption, then w can be expanded to a model of T_1 *and* it can be expanded to a model of T_2 , violating wod. Thus, we have that

Underdetermination of theories in the sense of (E-Indiscrim) implies *failure* of their weak-observational distinguishability.

This direction of Earman's reasoning is still true on our definitions. Underdetermination is surely a threat to applicability of the Gaifman-Snir Theorem. But it is a much harder condition to fulfill, if E-indiscriminability be the appropriate standard.

Still, as already suggested, there are many ways in which wod can fail short of genuine instances of underdetermination. Suppose a pair of theories agree—to within limits of observation, say—on all actual observations in a possible empirical world, w , but they clearly disagree in their predictions on the outcomes O of certain describable experiments Ex, *were they to be performed*. (For example, one theory might be classical Newtonian particle mechanics, the other special relativistic particle mechanics, as applied to a world of slowly moving particles, without any experiments to test the constancy of the velocity of light or other relativistic effects, etc.) We may suppose, further, that the statements corresponding to setting up and performing the experiments Ex can be given in the language of the evidence, as can the descriptions of possible outcomes of those experiments. In our experimentally impoverished world, w , the statement that an experiment of the class Ex is set up and performed is false, and so are all statements to the effect

that an outcome O is observed; all (material) conditionals to the effect that “if experiment of E_x is performed, O_i will result” are vacuously true, and so on. This is all perfectly compatible with *both* theories, even though they clearly disagree on *counterfactual* conditionals to the effect that “if E were performed, outcome O_i (say) *would have* resulted,” i.e., the theories are readily discriminable by *possible* evidence. Such locutions are not in the first-order formulations of the theories to which Gaifman-Snir applies; they can be gotten at only indirectly by looking to various models other than w . In any case, w clearly gives rise to a violation of wod, as it can be expanded to a model of the classical theory and it can be expanded to a model of the relativistic one. (This is so, provided the trajectories of particles in w are described not exactly but only to within limits accommodated by both theories.) The upshot is that, so long as our empirical or observational base is fairly restricted, applicability of Gaifman-Snir to theoretical hypotheses is severely limited indeed. The distinguishability condition needed (wod) is *easily* violated, without bringing in anything so radical as underdetermination of theories by possible evidence.

The only real hope, it seems to me, for restoring applicability of Gaifman-Snir is to expand the observation-base, allowing ever more sophisticated instrumentation and indirect methods of “observation” characteristic of actual scientific development. This recognizes that what counts as “theoretical” at a given stage may be counted as “observational” or “evidential” at a later stage. Thus, for example, the proposition that the mites the Chiharas describe (1994) *have legs* can—by the mid-twentieth century at least—be used as *evidence* for other statements, even though—as the light microscope is required for detection of the legs—it would be counted as “beyond the phenomena” in certain philosophies, and therefore a matter about which we must remain forever in the dark, as it were!

Recall that the definition of wod applies to a given partition of theories. The components of such a partition may themselves be disjunctions of theories. Presumably, on a narrow evidential base, those disjunctions must be rather broad in order to be wod from one another, i.e., major theoretical alternatives will remain unresolved even in the limit of accumulating evidence. (We may learn that “either the world is Newtonian or it is Einsteinian,” but nothing more.) Clearly as the observational base is expanded, the prospects are better that the theories of more and more refined partitions will satisfy wod. It is along these lines that partial successes in applying the merger-of-opinion results to theories may be achieved but how far one may go remains to be seen.

6. Concluding Remark: “But Can’t I Change My Mind?” An obvious

limitation of the merger-of-opinion results is their limiting character. As Earman brings out (in Ch. 6), information about rates of convergence is lacking, and probably unobtainable in the general setting. What this implies, for example, is that any finite number of arbitrary shifts in *Pr* functions is tolerated, so long as one remains within the class of equally dogmatic functions. Of course, most shifts would be regarded as the height of irrationality by virtually any scientific community, but the most that the merger results can say against them is that they may delay somewhat convergence on the truth—not much of a condemnation. Indeed, official Bayesianism pronounces illegal any shifts other than by conditionalization (strict or perhaps non-strict, e.g., of Jeffrey type), and it has sought to provide a rationale for such restrictions in the so-called “diachronic Dutch Book” arguments, to the effect that any shift other than by conditionalization can land the agent in the predicament of an inevitable loss. On the other side, one of Earman’s principal points is that certain “non-Bayesian shifts”—shifts that cannot be characterized as conditionalization in any reasonable sense—are inevitable in the progress of science, not only in revolutions but in the course of “normal science” as well, to use Kuhnian parlance. A prime example is the sort of shift that occurs upon exploration of novel, hitherto unimagined theoretical possibilities in a given domain, such as those Earman describes in connection with General Relativity. Moreover, such explorations, Earman argues, are often essential in mapping out a framework of possible alternatives so that a kind of “eliminative induction” can proceed. Bayesian methods are relevant in such a process, but they cannot be the whole story.

Many deep questions are raised here which cannot be taken up in this paper. Rather we shall conclude with a remark concerning the force of the diachronic Dutch Book arguments in light of the combination of Earman’s point and our construal of Dutch Book arguments above in Section 2.

The remark is this. Recall that, on the Howson and Urbach view and as we have redescribed them, Dutch Book arguments are really consistency arguments not involving decision or preference. In the diachronic setting, however, they have usually been cast as showing that, unless you conditionalize, you will be caught in a conflict with your former self, one which can be exploited to your inevitable detriment, e.g., by your being induced to sell back bets at what you now compute their fair value—assuming, that is, that you bought them in the first place. Much of the discussion of the force of such arguments has been carried on in a decision-theoretic context (cf. e.g., Maher 1992, as well as Earman’s discussion in his Ch. 2). But there is as much reason for distinguishing epistemic from decision-theoretic rationality in the dia-

chronic setting as there is in the synchronic. If we look at diachronic DB arguments purely from the standpoint of rationality_e, what do they show? Merely that shifts violating conditionalization commit you to judgments of fairness—better, of 0 expected flow—which are inconsistent with previous such judgments. (And conditionalization shifts themselves do not involve any such conflict, since a conditional probability $\text{Pr}(H/E)$ reflects only a *conditional* assessment of fairness, viz. a betting quotient on H that *would* be fair *were* the condition E taken as known. Thus, conditionalization can be seen as an application of *modus ponens* to certain conditionals (cf. Howson and Urbach 1989, 68).¹¹ But if we have altered our opinions about probabilities, including priors, what should we expect? *Of course*, our new opinions are in contradiction with some of our old ones—that’s what it is to change one’s mind! (Cf. Christensen 1991 and Howson 1992.) A Dutch Book construction to reach this conclusion is *recherché* to say the least. Moreover, unlike the (idealized) synchronic case, no obvious general conclusion concerning epistemic rationality can be drawn. If, in light of new information—especially new information not captured in sentences of our earlier language on which we could conditionalize—we assign new weights to some of our beliefs, we cannot be deterred merely by being reminded, “But this conflicts with what you said last week!” (The alert music student may confront his overbearing pedagogue with such a remark, only to endure as reply the quotation heading this section.) Even the fact, just acknowledged, that conditionalization can be cast as an application of *modus ponens* to certain counterfactuals does not imply that any such application is a requirement of rationality_e, for we may be well-advised to alter our previous *conditional* probability in light of intervening experience. *Modus ponens* has *two* premises, and in the diachronic setting they are indexed by different times; in between initial acceptance of the counterfactual conditional and the evidence statement E’s becoming known, a change in our assessment of what *would* be fair *were* E known can occur.¹² And if it is replied that the rule of conditionalization says, “Assign $\text{P}(H/E)$ to H when E and *only* E becomes known,” then nothing is said about the reasonableness of

11. More recently, however, Howson has taken a different view of conditionalization (expressed in a new paper, “Bayesian Rule of Updating,” forthcoming). Interestingly, Howson now comes to substantially the same conclusion as I do above on the impotence of dynamic Dutch Book arguments, and for essentially the same reasons.

12. Cf. Maher 1992, 136, who seems to be making this point. Also relevant is Levi’s (1980) distinction between diachronic conditionalization—at issue here—, and an atemporal notion he calls “confirmational conditionalization,” reflecting an agent’s confirmational commitments at a time concerning hypothetical belief states that could arise on learning new evidence.

shifting to $P'(H/E) \neq P(H/E)$ when something beyond E becomes known or experienced, even if one grants that sense can be made of “*only* E becomes known.”

Seen in this light, diachronic Dutch Book arguments do accomplish something: they highlight the distinction between conditionalization and categorical shifts involving conflict with earlier beliefs, and they provide a canonical way of expressing such conflicts (in terms of judgments of fairness or expected flow). But they do little more. They surely do not tell us that it is irrational_c to change our minds! (Of course, from the decision-theoretic standpoint, they might imply that, in certain cases, say, in which you have already purchased some bets, it is irrational_d both to change your mind and then to continue “to put your money where your mouth is,” e.g., by selling back the bets at a loss to protect against a newly perceived greater loss. But from the epistemic standpoint, such matters are entirely adventitious and irrelevant.)

No wonder it is necessary to go beyond Bayesianism. As powerful and useful a framework as it may be—and as, we hope, has been further illustrated in previous sections—the idea that all growth in knowledge can be captured in a small number of conditionalization rules comes perilously close to the doctrine of the *Meno*, that in a sense everything is already known in some prior condition. As useful as such a model may be for certain specific problems—e.g., application of a well-defined statistical model of the sort considered by van Fraassen (1989 163–170)—it cannot be all-encompassing. The argument for this conclusion that seems to me decisive—beyond merely pointing to lacunae in the justifications of conditionalization—is already present in Earman’s discussion of “non-Bayesian shifts” (see his Ch. 7), but can be sharpened: When new theoretical possibilities are explored, often *new concepts* are created and learned. These in turn give rise to *new propositions* S that simply could not be expressed in previous language. But the mere *possibility* that such propositions might be true can be enough to alter degrees of earlier held beliefs, i.e., we get a shift that the Bayesian should like to describe as passing from

$$\Pr(A/\diamond S), \text{ or } \Pr(A/\diamond S \ \& \ K),$$

where K is background knowledge, to

$$\Pr_{\diamond S}(A), \text{ or } \Pr_{\diamond S}(A/K),$$

i.e., adopting as new measure the conditional measure on $\diamond S$. But this model is simply inapplicable as $\diamond S$ is inexpressible in the old language. There is nothing around to conditionalize on until it is too late!¹³ Nor

13. For a recent effort to get around such difficulties raised by “new theories,” see

does it seem that Jeffrey conditionalization is generally applicable, for how are we to identify a partition expressible in the old language affected by the new knowledge that $\diamond S$? Learning of new theoretical possibilities such as that energy might be quantized or that light might be “partly a wave and partly a particle,” etc., is hardly analogous to dimly perceiving the color of some jelly beans! (Notice that the previously expressible, “Light is neither a wave nor a particle,” call it T_j , cannot serve if it was assigned 0 probability, since it then would contribute nothing to the right side of the Jeffrey scheme,

$$\Pr_{\text{new}}(A) = \sum_i \Pr_{\text{old}}(A/T_i) \times \Pr_{\text{new}}(T_i).$$

And even if one could identify a partition—say of theories previously believed possible—so that the Jeffrey conditionalization scheme could be applied, it must be remembered, as Earman pointed out, that the

Maher 1995. Two approaches are described. The second, which Maher favors, is set firmly within the framework of utility theory and relies on decisions “to accept or reject a theory” (p. 112); this approach is clearly unavailable to epistemic Bayesianism pursued in this essay. The first approach—one which Maher ultimately rejects as overly idealized—utilizes indirect references to theories yet to be formulated, e.g., by their ordering in a temporal sequence (A_{k+1}) (metalinguistically) stands for a singular term, say, ‘the next theory (in this field) to be formulated’), and it is argued that even after the content (expressed by the *proposition* B_{k+1}) of “the next theory” becomes available, updating of beliefs in propositions incorporating this content can be effected by conditionalizing on statements “about the new theory” that were previously given credibility values (e.g., “ A_{k+1} will have theoretical virtue v to degree m ”). The idea is that once the formulation B_{k+1} has been introduced, one has in effect learned the biconditional, “ A_{k+1} is true $\equiv B_{k+1}$ ”, and also the identity, “ A_{k+1} = the theory expressing that B_{k+1} ”. (In the interests of Maher’s proposal, I have replaced his single condition, “ $A_{k+1} = B_{k+1}$ ”, with these two to avoid a serious use/mention confusion: B_{k+1} is a proposition in the new B-algebra, not a singular term.) This permits substituting the left sides for the right, respectively, throughout the new B-algebra of propositions preserving the new q-probabilities. One has thereby passed back to the old A-algebra on which the old p-probabilities are defined. So *if* probabilities on the A-algebra are updated by conditionalization, then so are probabilities on the B-algebra. However, it would be question-begging to *assume* that “the probabilities on the A-algebra are updated by conditionalization” (as Maher requires, p. 108). Suppose B_{k+1} opens up an unforeseen possibility not previously expressible which immediately leads me to reduce the credibility of a previously considered theory, T, which prior to formulation B_{k+1} I predicted would be unaffected by introduction of “the next theory”; then $q(T/A_{k+1} \text{ is introduced}) = q(T/\text{“the theory expressing that } B_{k+1} \text{ is introduced”}) < p(T) = p(T/A_{k+1} \text{ is introduced})$ (where reference to further background knowledge has been suppressed throughout), and there may be nothing further to conditionalize on that would make any difference. In fairness, Maher is only claiming that there is a logical possibility in an abstract setting of there being something further to conditionalize on (in order to refute a certain too sweeping general argument), and that can be conceded. But that is hardly a solution to Earman’s problem, as Maher appears to recognize, or to the related ones highlighted in this paper.

merger-of-opinion results depend on strict conditionalization. This is not surprising, for what is to constrain the $\text{Pr}_{\text{new}}(T_i)$?

Finally, it should be noted that similar remarks on the limits of Bayesian learning apply also in connection with *observations* of new phenomena, whose scientific description may require new language. Jeffrey conditionalization was introduced to treat cases in which one has some sensory experience not adequately described by existing “observation terms.” But we should also consider observations of novel phenomena in science that come to be described in what was previously taken as “theoretical vocabulary,” e.g., early observations of atomic radiation, or of spatial quantization associated with quantum mechanical spin, etc. Admittedly, according to traditional versions of empiricism, new statements such as “Electron spin is (or is here behaving as) quantized,” or “RNA is present at the ribosomes of these cells,” count as permanently “theoretical,” unlike statements referring only to darkened spots on a plate, etc. If observations are to be described in terms of a fixed “observation language,” phenomenal or macro-physical, then there is normally no problem with conditionalization due to expressive incompleteness. But in these terms, as we have seen above, there also can be little hope of applying the merger-of-opinion results to theoretical hypotheses due to problems of distinguishability. Thus, there seems to be a trade-off: if we broaden the “observation base” in the interests of distinguishability for purposes of applying the merger results, we must also countenance more gaps in the learning models based on conditionalization. If, at a given stage, we allow statements such as “Electron spin is (behaving as) quantized,” or “RNA is present at these ribosomes,” etc., to count as evidence for further hypothesis testing, we may achieve greater distinguishability among hypotheses; but, if the stage is relatively early—soon after introduction of the new language—then conditionalization on such “evidence” becomes problematic since the needed “old” conditional probabilities did not yet exist. (This can be termed “the problem of *new* evidence”—a companion of Earman’s “problem of new theories” (Ch. 5, §6).) But this seems a direction worth pursuing. As Earman’s considerations suggest, there are gaps enough already in Bayesian learning models; and, in any case, gaps are not errors; “much needed” or not, they can sometimes be filled.¹⁴

14. Recall the quip about the research paper that “filled a much needed gap in the literature”!

APPENDIX

Here we prove the lemma of Section 3, which we first restate:

- (1) $\Pr(A/N \ \& \ \tilde{E}) = 0$
 (2) $\Pr(K) = \Pr(K/N) = 1$
 (3) $\varepsilon_N = \text{df } \Pr(K) - \Pr(N \ \& \ K)$

LEMMA: (1) – (3) $\Rightarrow \Pr(A \ \& \ \tilde{E} \ \& \ K) \leq \varepsilon_N$.

Proof: As an aid to reasoning, let us use set-theoretic Boolean operations. Translation back to sentential logical notation is routine. First, some elementary set-theoretic reminders:

$$(i) (X \cap Z) - (Y \cap Z) \subseteq X - Y$$

Symmetric difference is defined by

$$X \Delta Y = \text{df } (X - Y) \cup (Y - X).$$

From (i),

$$(ii) (X \cap Z) \Delta (Y \cap Z) \subseteq X \Delta Y.$$

Also, recall

$$(iii) X - Y = X - (X \cap Y).$$

So, for any measure m , by disjoint additivity,

$$(iv) m(X - Y) = m(X) - m(X \cap Y).$$

Now set $\varepsilon = \Pr(N) - \Pr(N \cap K)$. By (2), $\varepsilon = 0$. Then

$$\Pr(N \Delta K) = \Pr([N - (N \cap K)] \cup [K - (K \cap N)]) = \Pr(N - (N \cap K)) + \Pr(K - (K \cap N)) = \varepsilon + \varepsilon_N = \varepsilon_N.$$

By (ii),

$$\Pr((A \cap \tilde{E} \cap N) \Delta (A \cap \tilde{E} \ \& \ K)) \leq \varepsilon_N,$$

and by (1) and (iv), therefore,

$$\Pr(A \cap \tilde{E} \cap K) - \Pr(A \cap \tilde{E} \cap N \cap K) = \Pr(A \cap \tilde{E} \cap K) \leq \varepsilon_N,$$

which is the desired result.

REFERENCES

- Cartwright, N. (1979), "Causal Laws and Effective Strategies", *Nous* 13: 419–437.
 Chihara, C. and C. Chihara (1993), "A Biological Objection to Constructive Empiricism", *British Journal for the Philosophy of Science* 44: 653–658.
 Christensen, D. (1991), "Clever Bookies and Coherent Beliefs", *Philosophical Review* 50: 229–247.
 Earman, J. (1992), *Bayes or Bust?* Cambridge, MA: MIT Press.
 Fernie, J. D. (1995), "The Neptune Affair", *American Scientist* 83, March–April: 116–119.

- Friedman, M. (1974), "Explanation and Scientific Understanding", *Journal of Philosophy* 70: 5–19.
- Gardner, M. R. (1983), "Realism and Instrumentalism in Pre-Newtonian Astronomy", in J. Earman (ed.), *Testing Scientific Theories, Minnesota Studies in the Philosophy of Science, Vol. X*. Minneapolis: University of Minnesota Press, 1983, pp. 201–265.
- Hellman, G. and F. Thompson (1975), "Physicalism: Ontology, Determination, and Reduction", *Journal of Philosophy* 71: 551–564.
- Horwich, P. (1982), *Probability and Evidence*. Cambridge: Cambridge University Press.
- Howson, C. (1992), "Dutch Book Arguments and Consistency", in D. Hull, M. Forbes, and K. Okruhlik (eds.), *PSA 1992, Vol. Two*. East Lansing, MI: Philosophy of Science Association, 1993, pp. 161–168.
- . (forthcoming) "Bayesian Rules of Updating", *Erkenntniss*.
- Howson, C. and P. Urbach (1989), *Scientific Reasoning: The Bayesian Approach*. La Salle, IL: Open Court.
- Kim, J. (1984), "Concepts of Supervenience", *Philosophy and Phenomenological Research* 65: 153–176.
- Kitcher, P. (1989), "Explanatory Unification and the Causal Structure of the World", in P. Kitcher and W. Salmon (eds.), *Scientific Explanation: Minnesota Studies in the Philosophy of Science, Vol. 13*. Minneapolis: University of Minnesota Press, 1989, pp. 410–505.
- Kyburg, H. E., Jr. (1983), *Epistemology and Inference*. Minneapolis: University of Minnesota Press.
- Maher, P. (1992), "Diachronic Rationality", *Philosophy of Science* 59: 120–141.
- . (1995), "Probabilities for New Theories", *Philosophical Studies* 77: 103–115.
- Mayo, D. (1991), "Novel Evidence and Severe Tests", *Philosophy of Science* 58: 523–552.
- Pollock, J. L. (1986), *Contemporary Theories of Knowledge*. Savage, MD: Rowman & Littlefield.
- Rosenkranz, R. (1977), *Inference, Method, and Decision*. Dordrecht: Reidel.
- Schick, F. (1986), "Dutch Bookies and Money Pumps", *Journal of Philosophy* 83: 112–119.
- van Fraassen, B. C. (1980) *The Scientific Image*. Oxford: Oxford University Press.
- . (1989) *Laws and Symmetries*. Oxford: Oxford University Press.
- Wayne, A. (1995) "Bayesianism and Diverse Evidence", *Philosophy of Science* 62: 111–121.