**Bayes or Bust?**
**A Critical Examination of Bayesian Confirmation Theory**

John Earman

The reader who penetrates very far into this work may begin to experience a topsy-turvy feeling. Those who complain will in effect be crying, Stop the world, I want to get off! For the world of confirmation is a topsy-turvy world.

My intellectual debts are too numerous to recount here. But I would be remiss if I did not acknowledge the many helpful suggestions I received on earlier drafts of this book. Thank you Jeremy Butterfield, Charles Chihara, Alan Franklin, Donald Gillies, Clark Glymour, David Hillman, Colin Howson, Richard Jeffrey, Cory Juhl, Kevin Kelly, Philip Kitcher, Tim Maudlin, John Norton, Teddy Seidenfeld, Elliott Sober, Paul Teller, Jan von Plato, Bas van Fraassen, and Sandy Zabell. With all of this help, the book should be better than it is.

The material in chapter 1 appeared as "Bayes' Bayesianism," *Studies in the History and Philosophy of Science* 21 (1990): 351–370. A version of chapter 5 appeared as "Old Evidence, New Theories: Two Unsolved Problems in Bayesian Confirmation Theory," *Pacific Philosophical Quarterly* 70 (1989): 323–340. I am grateful to the editors of these journals for their permission to reprint this material. The excerpts from Bayes's essay found in the appendix to chapter 1 are reproduced with the kind permission of the editors of *Biometrika*.

# Notation

## Logic

| | |
|---|---|
| $(\forall x)$ | universal quantifier |
| $(\exists x)$ | existential quantifier |
| $\&$ | conjunction, $A \mathbin{\&} B$ ($A$ and $B$) |
| $\&_{i \leq n} A_i$ | $A_1 \mathbin{\&} A_2 \mathbin{\&} \ldots \mathbin{\&} A_n$ |
| $\vee$ | disjunction, $A \vee B$ ($A$ or $B$ or both) |
| $\bigvee_{i \leq n} A_i$ | $A_1 \vee A_2 \vee \ldots \vee A_n$ |
| $\rightarrow$ | material implication, $A \rightarrow B$ (if $A$ then $B$) |
| $\leftrightarrow$ | biconditional, $A \leftrightarrow B$ ($A$ if and only if $B$) |
| $\neg$ | negation, $\neg A$ (not $A$) |
| $\equiv$ | definition |
| $\models A$ | $A$ is valid (i.e., $A$ is true in every possible world or model) |
| $A \models B$ | $A$ semantically implies $B$ (i.e., $B$ is true in every possible world or model in which $A$ is true) |

## Mathematics

| | |
|---|---|
| $x \in X$ | $x$ is an element of $X$ |
| $X \subseteq Y$ | $X$ is a subset of $Y$ |
| $X \cap Y$ | the intersection of $X$ and $Y$ |
| $X \cup Y$ | the union of $X$ and $Y$ |
| $\sum_i$ | summation over the index $i$ |
| $\int$ | integral |
| $\mathbb{N}$ | natural numbers |
| $\mathbb{R}$ | real line |
| $f: X \rightarrow Y$ | $f$ maps $x \in X$ to $f(x) \in Y$ |
| sup | supremum |
| $\lim_{n \to \infty}$ | the limit as $n$ approaches infinity |

**Probability and Statistics**

$\Pr(A)$      the (unconditional) probability of $A$

$\Pr(A/B)$      the conditional probability of $A$ on $B$

$E(X)$      the expectation value of the random variable $X$

$\binom{n}{m}$      $n!/[m!(n-m)!]$

# 5 The Problem of Old Evidence

## 1 Old Evidence as a Challenge to Bayesian Confirmation Theory

One of the great virtues of Bayesian confirmation theory is its ability to pinpoint and explain the strengths and weaknesses of rival accounts, or so it was claimed in chapters 3 and 4. Recall, in particular, that I claimed that Bayesianism explains why the HD method of confirmation works as well as it does. To review, suppose that the confirmatory power of $E$ for $T$ is measured by $C(T, E) \equiv \Pr(T/E) - \Pr(T)$.[1] Suppose further that (1) $T \models E$ (the basic HD condition), (2) $1 > \Pr(T) > 0$, and (3) $1 > \Pr(E) > 0$. Then by Bayes's theorem, $C(T, E) > 0$, in consonance with the HD method. Furthermore,

$$C(T \ \& \ X, E)/C(T, E) = \Pr(T \ \& \ X)/\Pr(T) = \Pr(X/T),$$

which shows how tacking an irrelevant conjunct $X$ onto $T$ serves to reduce confirmatory power.

This display of virtue also serves to reveal an Achilles' heel of Bayesianism, or so Glymour (1980) has argued. To see the difficulty in concrete terms, take the time to be November 1915, when Einstein formulated the final version of his general theory of relativity (GTR) and when he first showed that this theory explains the heretofore anomalous advance of the perihelion of Mercury.[2] The nature of Mercury's perihelion had been the subject of intensive study by Le Verrier, Newcomb, and other astronomers, and so the relevant facts $E$ were old evidence. In Bayesian terms, this means that for any agent who was conversant with the field and who operated according to the model of learning from experience by strict conditionalization (see chapter 2), $\Pr_{1915}(E) = 1$. Thus condition (3) above fails, with the results that $\Pr_{1915}(T/E) = \Pr_{1915}(T)$ and $C(T, E) = 0$ for any $T$ and thus for GTR in particular, which seems to run counter to the generally accepted conclusion that the facts $E$ did in November 1915 (and still do) provide strong confirmation of GTR. Indeed, in an exhaustive survey of the literature Brush (1989) found that with very few exceptions physicists have held that the perihelion phenomenon gives better confirmational value than either of the other two classical tests—the bending of light and the red shift—despite the fact that the former was old news while the latter two represented novel predictions.

Replacing the incremental with the absolute criterion of confirmation, according to which $E$ confirms $T$ at $t$ just in case $\Pr_t(T/E) > r$ for some fixed $r > 0$, allows old evidence to have confirmational value. But the value

of the absolute degree of confirmation may not capture the strength of the evidence, since if $E'$ and $E''$ are both old at $t$, then $\text{Pr}_t(T/E') = \text{Pr}_t(T/E'') = \text{Pr}_t(T).$[3]

In section 2, I will examine some preliminary attempts to solve or dissolve the problem of old evidence. Although unconvincing, these attempts nevertheless serve the useful functions of distinguishing different versions of the problem and of pinpointing the version most worthy of attention. Sections 3 to 5 discuss an approach to this worthy version due to Garber (1983), Jeffrey (1983a), and Niiniluoto (1983). Section 6 takes up the neglected flip side of the old evidence problem: the problem of new theories. Section 7 summarizes some pessimistic conclusions for the prospects of Bayesian confirmation theory in the light of the old-evidence problem.

## 2   Preliminary Attempts to Solve or Dissolve the Old-Evidence Problem

Resorting to an objectivist interpretation of probability would help if it led to $\text{Pr}(E) < 1$ for old evidence $E$. But insofar as I have a grasp of objective probability interpreted either as propensity or as relative frequency, it would seem that the objective probability is 1 (or else 0) for an anomalous perihelion advance of 43 seconds of arc per century, at least on the assumption that a deterministic mechanism is operating. This difficulty need not arise if objective probability means not propensity or frequency but the uniquely determined rational degree of belief. Thus Rosenkrantz (1983) recommends that we compute $\text{Pr}(E)$ relative to a "considered partition" $H_1, H_2, \ldots, H_n$: $\text{Pr}(E) = \sum_{i=1}^{n} \text{Pr}(E/H_i) \times \text{Pr}(H_i)$. He claims that unless $E$ is a necessary truth, this sum will be less than 1 and will remain less than 1, since the likelihoods $\text{Pr}(E/H_i)$ are "timeless relations." I have two difficulties with this tack. First, I am unpersuaded by attempts to objectify the assignments of prior probabilities.[4] Second, if $\text{Pr}(\cdot)$ is interpreted as degree of belief, rational or otherwise, then it must be time-indexed, and $\text{Pr}_{1915}(E)$ would seem to be 1.

Another way to try to resolve the problem of old evidence would be to insist upon using a conditional probability function $\text{Pr}(\cdot/\cdot)$ that is defined even when the conditioning propositions have 0 unconditional probability (see appendix 1 to chapter 2). This allows one to adopt as a measure of evidential support $\hat{C}(T, E) \equiv \text{Pr}(T/E) - \text{Pr}(T/\neg E)$. When $0 > \text{Pr}(T) > 1$,

$C(T, E) = (1 - \text{Pr}(T))\hat{C}(T, E)$, so the new measure agrees qualitatively with the old as to positive and negative relevance. But the seeming advantage of the new measure is that in the case of old evidence ($\text{Pr}(E) = 1$), when $C(T, E) = 0$, $\hat{C}(T, E)$ can be positive. Unfortunately, $\hat{C}(T, E)$ is not a suitable measure of confirmatory power. In the HD case ($T \models E$) with old evidence ($\text{Pr}(E) = 1$), $\hat{C}(T, E) = \text{Pr}(T)$, which means that all such evidence is counted as having the same confirmatory power.[5]

The original problem of old evidence would vanish for Bayesian personalists for whom $\text{Pr}(E) \neq 1$, with Pr interpreted as personal degree of belief.[6] There are both historical and philosophical reasons for such a stance. In my running example, the literature of the period contained everything from $41''$ to $45''$ of arc per century as the value of the anomalous advance of Mercury's perihelion, and even the weaker proposition that the true value lies somewhere in this range was challenged by some astronomers and physicists.[7] Of course, if we push this line to its logical conclusion, we will eventually reach the position that no "thing language" proposition of the sort useful in confirming scientific theories is ever learned for certain, and the strict conditionalization model will collapse. Bayesians are hardly at a loss here, since Jeffrey (1983b) has proposed a replacement for strict conditionalization that allows for uncertain learning (see chapter 2).

However, denying that $\text{Pr}(E) = 1$ only serves to trade one version of the old-evidence problem for another. Perhaps it was not certain in November 1915 that the true value of the anomalous advance was roughly $43''$ of arc per century, but most members of the scientific community were pretty darn sure, e.g., $\text{Pr}(E) = .999$. Assuming that Einstein's theory does entail $E$, we find that the confirmatory power $C(T, E)$ of $E$ is $\text{Pr}(T) \times .001/.999$, which is less than .001002. This is counterintuitive, since, to repeat, we want to say that the perihelion phenomenon did (and does) lend strong support to Einstein's theory.[8]

In what follows, then, I will work within the Bayesian personalist framework, using strict conditionalization to model learning from experience, and I will use $C(T, E)$ as the measure of confirmatory power. I will first consider the response that a proper use of this apparatus shows the old-evidence problem to be a pseudoproblem for logically omniscient Bayesian agents. Here logical omniscience involves two elements. The first (LO1) embodies the assumption that all the logical truths of the language $L$ on which Pr is defined are transparent to the Bayesian agent. This assumption is codified in the basic axiom that if $\models X$ in $L$, then $\text{Pr}(X) = 1$. Thus failure to accord maximal probability to logical truths of $L$ leads to Dutch-book

situations. The second element (LO2) involves the assumption that the agent is aware of every theory that belongs to the space of possibilities. In effect, when making the starting probability assignments $\text{Pr}_{t_0}$, the agent formulates and considers every theory that can be stated in $L$. Now take a piece of empirical evidence $E$ about which the agent was not certain *ab initio*, i.e., $1 > \text{Pr}_{t_0}(E) > 0$, and suppose that the agent learns $E$ between $t_n$ and $t_{n+1}$. Then if $\text{Pr}_{t_n}(T/E) > \text{Pr}_{t_n}(T)$, a confirmational event takes place.[9] That event takes place only once, since on the strict conditionalization model, for any $m > n + 1$, $\text{Pr}_{t_m}(E) = 1$. But once is enough, for at $t_m$ we can still say that $E$ is good evidence for $T$, since the history of the present probability function $\text{Pr}_{t_m}$ contains the relevant sort of confirmational event.

This solution does not apply to real-world Bayesian agents who violate (LO2). In my running example, this includes the entire physics community in 1915, since Einstein's general theory was not formulated until the end of November of that year. Of course, if we could succeed in showing in Bayesian terms how GTR was confirmed for real-life scientists in 1915, then we could use the above strategy to cover post-1915 times.

As an aside it may be helpful here to refer to Eells's (1985) revealing classification of the problems of old evidence:

I. The problem of *old new evidence*: $T$ was formulated before the discovery of $E$, but it is now later, and $\text{Pr}(E) = 1$. So $\text{Pr}(T/E) = \text{Pr}(T)$.

II. The problem of *old evidence*: $E$ was known before the formulation of $T$.
   A.  The problem of *old old evidence*: It is now some time subsequent to the formulation of $T$.
       1.  $T$ was originally designed to explain $E$.
       2.  $T$ was not originally designed to explain $E$.
   B.  The problem of *new old evidence*: It is now the time (or barely after the time) of the formulation of $T$.
       1.  $T$ was originally designed to explain $E$.
       2.  $T$ was not originally designed to explain $E$.

Taking into account confirmational histories and confirmational events seems to solve (I), and given a solution to (II.B), it also serves to solve (II.A). The remaining problem is that of new old evidence. The only places I differ with Eells are over cases (II.A.1) and (II.B.1), where Eells assumes that $E$ cannot confirm $T$ (see the discussion below in sections 5 and 6).

To return to the main discussion, Garber (1983), Jeffrey (1983a), and Niiniluoto (1983) have reacted to this version of the old-evidence problem by proposing to drop (LO1) as well as (LO2). Dropping (LO1) allows Bayesian agents to do logical and mathematical learning, and such learning, so they claim, can serve to boost the probability of the theory, as required by the incremental analysis of confirmation. What Einstein learned in November of 1915, so the story goes, was that his general theory entailed the heretofore anomalous perihelion advance, and conditionalization on that new knowledge was the relevant confirmational event.[10] I will examine this line of attack in detail in the following sections, but before doing so, I will comment briefly on another tack.

The alternative to Garber, Jeffrey, and Niiniluoto (GJN) is to demonstrate incremental confirmation for counterfactual degrees of belief, using degrees of belief the agent would have had if he hadn't known $E$ prior to the formulation of $T$.[11] But as Chihara (1987), Eells (1985), Glymour (1980), van Fraassen (1988) and other commentators have objected, it is not evident that the relevant counterfactual degree of belief will be determinate or even that it will exist. In my historical example it is relevant that in 1907 Einstein wrote, "I am busy on a relativistic theory of the gravitational law with which I hope to account for the still unexplained secular change of the perihelion motion of Mercury. So far I have not managed to succeed" (Seelig 1956, p. 76). Thus it is not beyond the pale of plausibility that if Einstein hadn't known about the perihelion phenomenon, he wouldn't have formulated GTR. And if someone else had formulated the theory, Einstein might not have taken it seriously enough to assign it a nonzero prior, or he might not have understood it well enough to assign it any degree of belief at all. I will return to counterfactual degrees of belief in section 7.

## 3  Garber's Approach

To illustrate how logical omniscience (LO1) can be abandoned so as to make way for a more realistic Bayesianism, Garber (1983) begins with a language $L$ in which distinct atomic sentences are treated as logically independent and in which the nonatomic sentences are all truth-functional compounds of the atomic ones. He then moves to a richer language $L^*$ that contains the sentences of $L$ and also new atomic sentences of the form $X \vdash Y$, where $X$ and $Y$ are sentences of $L$.[12] The symbol '$\vdash$' is a primitive

connective of $L^*$, but the aim is to interpret it as logicomathematical implication in whatever system of logic and mathematics is needed for the branch of science in question. Toward this end, Garber requires that under the Pr function, '$\vdash$' behaves as if it obeys modus ponens:

$$\Pr((X \vdash Y) \,\&\, X) = \Pr((X \vdash Y) \,\&\, X \,\&\, Y) \tag{G}$$

Garber then shows that learning that $T \vdash E$ can serve to confirm $T$. More specifically, he shows that there is a probability Pr defined on the sentences of $L^*$ and satisfying (G) such that $0 < \Pr(A \vdash B) < 1$ whenever $A$ and $\neg B$ are not both tautologies of $L$. It may therefore be that $\Pr(A/A \vdash B) > \Pr(A)$. We may wish to add the further constraint that $\Pr(A \vdash B) = 1$ whenever $A \to B$ is a tautology of $L$. But since Einstein's GTR does not truth-functionally entail the perihelion advance evidence $E$, it is consistent with this constraint to set $\Pr(\text{GTR} \vdash E) < 1$ and thus to have $\Pr(\text{GTR}/\text{GTR} \vdash E) > \Pr(\text{GTR})$.

Three criticisms have been brought against this approach. The first is that Garber has only shown that a solution to the problem of old evidence is possible within the framework of the Bayesian strict conditionalization model and not that a solution of this form actually applies to the historical case at issue. To complete the solution for the case of the perihelion of Mercury, it would have to be demonstrated that there is a plausible set of constraints that Einstein's degrees of belief did or should have satisfied and that guarantee that his learning that $\text{GTR} \vdash E$ served to boost his degree of belief in GTR. This hiatus was addressed by Jeffrey (1983a), whose attempt to fill the gap will be examined in detail in the next section.

A second criticism of Garber's approach derives from the observation that his approach requires logical omniscience (LO1) with respect to the truth-functional logic of $L^*$ but not with respect to predicate logic, arithmetic, or calculus. But demanding knowledge of very complicated truth functional implications can be even more unrealistic than demanding knowledge of simple truths of arithmetic or calculus. This leads Eells (1985) to propose that a truly realistic version of Bayesianism should set the standard of what logicomathematical truths the Bayesian agent is supposed to know in terms of complexity. I agree with Eells, but for present purposes it suffices to stick to Garber's preliminary version of (not thoroughly) humanized Bayesianism.

A third, and I believe unfair, category of criticism has been leveled in both the published literature and in informal discussions. To put it at its

unfairest, the charge would go thus. To apply Garber's formalism to the problem of old evidence assumes that '$\vdash$' has been identified as the appropriate form of logicomathematical implication. But constraint (G) does not suffice for this identification, since other relations also satisfy (G). And if conditions sufficient to pin down the intended interpretation of '$\vdash$' are added to (G), there is no guarantee that Garber's demonstration of how learning $T \vdash E$ can serve to boost the probability of $T$ will remain valid.

The response I propose on Garber's behalf is that (G) is not supposed to fix the interpretation of '$\vdash$'. The interpretation is fixed extrasystematically, e.g., by the intention of the agent to use '$\vdash$' to mean implication in some logicomathematical system. If it is then asked why (G) was imposed in the first place, two reasons can be given. It is important to demonstrate that Pr assignments can be made to reflect various constraints that '$\vdash$' ought to satisfy if it is interpreted extrasystematically as logicomathematical implication. And also, (G) plays an important role in proving that in actual historical cases $\Pr(T/T \vdash E) > \Pr(T)$, as will be seen below in section 4.

Van Fraassen (1988) is unconvinced that Pr assignments that resolve the old evidence problem can be made to conform to constraints appropriate to '$\vdash$' taken as a form of logicomathematical implication. In particular, he proposes that the form of a plausible constraint for conditional proof is the following:

If $\Pr'(X \,\&\, Y) = \Pr'(X)$ for all $\Pr'$ in $\mathscr{C}(\Pr)$, then $\Pr(X \vdash Y) = 1$. (vF)

Here $\mathscr{C}(\Pr)$ is the class of alternatives to Pr that allow for the generalizations involved in conditional proof. Van Fraassen then poses the following dilemma for Garber. Suppose first that $\mathscr{C}(\Pr)$ consists of all probability functions that can be generated from Pr by strict conditionalization or by Jeffrey conditionalization or, more generally, by any shift in degrees of belief that does not change zero probabilities into nonzero probabilities. (In the jargon of probability, $\Pr'$ must be absolutely continuous with respect to Pr.) Then it follows from (vF) that

$$\Pr(X \vdash Y) = \Pr((X \vdash Y) \,\&\, (X \to Y))$$

and that

$$\Pr(X \to Y) = \Pr((X \to Y) \,\&\, (X \vdash Y)).$$

So $X \vdash Y$ is probabilistically indistinguishable from material implication. Thus if $Y$ is old news, then so is $X \vdash Y$. On the other hand, to take $\mathscr{C}(\Pr)$

to include probability functions not absolutely continuous with respect to Pr is to consider priors we might have had if we didn't have the old evidence. But that is to enter the mire of counterfactual beliefs that I decided above in section 2 must be avoided.

On behalf of Garber, I propose to escape the dilemma by rejecting (vF) as a suitable means of reflecting conditional proof. Indeed, one can hold that no condition like (vF) is needed to reflect the successful application of a proof strategy, whether conditional proof or otherwise. Rather, the success is expressed in the Bayesian learning model. Thus, suppose that at time $t_n$ the Bayesian agent shows that $X$ implies $Y$ by means of a conditional proof strategy, by a *reductio* strategy, by deriving $Y$ from $X$ and the accepted axioms by means of the accepted inference rules, or by whatever proof strategy is allowed in the relevant logic. Then the agent has learned that $X \vdash Y$, and so on the strict conditionalization model, $\mathrm{Pr}_{t_{n+1}}(X \vdash Y) = 1$.

My own concern about Garber's system lies not so much with qualms about lurking inconsistencies as with doubts about its relevance to the old-evidence problem. The axiom of probability requiring that $\mathrm{Pr}(X) = 1$ if $\models X$ is not contradicted in Garber's system by a value for $\mathrm{Pr}(\mathrm{GTR} \vdash E)$ lying strictly between 0 and 1, since in this system neither $\models (\mathrm{GTR} \vdash E)$ nor $\models \neg(\mathrm{GTR} \vdash E)$. For in Garber's system a possible world is given by an assignment of truth values to the atomic sentences; thus $\mathrm{GTR} \vdash E$ is true in some of the possible worlds and false in others. But then it is hard to see how the formal result that $\mathrm{Pr}(\mathrm{GTR}/\mathrm{GTR} \vdash E) > \mathrm{Pr}(\mathrm{GTR})$ bears on the idea that my learning that the theory entails $E$ (in predicate logic or second-order logic or whatever) boosts my degree of belief in GTR, for the formal result holds only for a semantics that masks what $\mathrm{GTR} \vdash E$ is supposed to mean. In what follows, I waive this qualm because I think that the entire approach is beset by more fundamental difficulties.

## 4  Jeffrey's Demonstration

Let us now consider Jeffrey's attempt to show how learning that $T \vdash E$ can serve to boost the probability of $T$. Suppose the following:

$$\mathrm{Pr}(E) = 1 \tag{J1.a}$$

$$1 > \mathrm{Pr}(T) > 0 \tag{J1.b}$$

$$1 > \mathrm{Pr}(T \vdash E) > 0, \qquad 1 > \mathrm{Pr}(T \vdash \neg E) > 0 \tag{J2.a}$$

$$\mathrm{Pr}((T \vdash E) \,\&\, (T \vdash \neg E)) = 0 \tag{J2.b}$$

$$\mathrm{Pr}(T/(T \vdash E) \vee (T \vdash \neg E)) \geqslant \mathrm{Pr}(T) \tag{J3}$$

$$\mathrm{Pr}(T \,\&\, (T \vdash \neg E)) = \mathrm{Pr}(T \,\&\, (T \vdash \neg E) \,\&\, \neg E) \tag{J4}$$

Then $\mathrm{Pr}(T/T \vdash E) > \mathrm{Pr}(T)$.

*Proof*

$$\mathrm{Pr}(T/(T \vdash E) \vee (T \vdash \neg E))$$

$$= \frac{\mathrm{Pr}(T \,\&\, (T \vdash E)) + \mathrm{Pr}(T \,\&\, (T \vdash \neg E)) - \mathrm{Pr}(T \,\&\, (T \vdash E) \,\&\, (T \vdash \neg E))}{\mathrm{Pr}(T \vdash E) + \mathrm{Pr}(T \vdash \neg E) - \mathrm{Pr}((T \vdash E) \,\&\, (T \vdash \neg E))}$$

$$= \frac{\mathrm{Pr}(T/T \vdash E)}{1 + [\mathrm{Pr}(T \vdash \neg E)/\mathrm{Pr}(T \vdash E)]}.$$

The first equality follows by the definition of conditional probability and the standard axioms of probability. The second equality follows from the first, since by (J1.a) and (J.4), the second and third terms in the numerator on the right-hand side are 0, and by (J2.b), the third term in the denominator is 0. Then by (J2.a), the right hand side of the second equality is less than $\mathrm{Pr}(T/T \vdash E)$. But by (J3), the right-hand side of the second equality is greater than or equal to $\mathrm{Pr}(T)$, which gives the desired result.

Conditions (J1) and (J2.a) follow from the meaning of the problem of old evidence. Condition (J4) is just an application of Garber's (G). The crucial condition is (J3), which says in application to my running example that in November of 1915 Einstein's degree of belief in GTR before learning that it entailed the missing 43″ was less than or equal to his conditional degree of belief in the theory, given that the theory implies a definite result about the perihelion advance. Eells (1985) demurs that (J3) is suspect. For example, the above demonstration shows that in the presence of the other conditions (J3) leads to the result that if $\mathrm{Pr}(T \vdash E) = \mathrm{Pr}(T \vdash \neg E)$, then $\mathrm{Pr}(T/T \vdash E) \geqslant 2\mathrm{Pr}(T)$. This means that the prior probability of $T$ cannot be greater than .5. And as $\mathrm{Pr}(T)$ approaches .5, $\mathrm{Pr}(T/T \vdash E)$ approaches 1, a wholly implausible result in the actual historical case at issue. I would add that (J2.b) is also suspect, for if we are supposed to be

imagining humanized, nonlogically omniscient agents, it is unreasonable for them to be certain that a new and complicated theory is internally consistent.

Such difficulties do not lead Eells to reject the GJN approach. His stance is that the Bayesian can explicate "$T \vdash E$ confirms $T$" as $\Pr(T/T \vdash E) > \Pr(T)$ "without expecting there to be any single formal kind of justification ... for exactly the cases in which $T \vdash E$ *should* be taken as confirming $T$" (1985, p. 299). While it is a fair comment that a single formal justification cannot be expected to cover all the cases, the GJN approach loses its interest if it cannot be shown that increases in probability will take place in an interesting range of cases. (Suppose that in the original setting, where the problem of old evidence is neglected, all the Bayesians could say is that $\Pr(T/E) > \Pr(T)$ happens when it happens. Then I would suggest that the number of adherents to Bayesian confirmation theory would dwindle. Fortunately, we can demonstrate that the inequality holds when various relations between $T$ and $E$ obtain and that these relations cover many of the cases where confirmation intuitively takes place.) It is to this question that I now turn.

As an alternative to Jeffrey's demonstration, consider the following:

$$\Pr(E) = 1 \tag{A1.a}$$

$$1 > \Pr(T) > 0 \tag{A1.b}$$

$$1 > \Pr(T \vdash E) > 0 \tag{A2}$$

$$\Pr((T \vdash E) \vee (T \vdash \neg E)) = 1 \tag{A3}$$

$$\Pr(T \& (T \vdash \neg E)) = \Pr(T \& (T \vdash \neg E) \& \neg E) \tag{A4}$$

Then $\Pr(T/T \vdash E) > \Pr(T)$.

*Proof*

$$\Pr(T) = \Pr(T \& [(T \vdash E) \vee (T \vdash \neg E)])$$

$$= \Pr(T \& (T \vdash E)) + \Pr(T \& (T \vdash \neg E))$$

$$\quad - \Pr(T \& (T \vdash E) \& (T \vdash \neg E))$$

$$= \Pr(T \& (T \vdash E)).$$

The first equality holds in virtue of (A3). The second follows from the first by the addition axiom. And the third follows in virtue of (A1.a) and (A4). Thus

$$\Pr(T/T \vdash E) - \Pr(T) = \frac{\Pr(T)}{\Pr(T \vdash E)} [1 - \Pr(T \vdash E)],$$

which, by (A1.b) and (A2), is greater than 0.

Since (A3) is stronger than Jeffrey's (J3), it might seem that this second derivation cannot be an improvement on the first. Note, however, that the present derivation did not have to rely on the suspect assumption that

$$\Pr((T \vdash E) \& (T \vdash \neg E)) = 0.$$

Yet it would also seem that this approach is also subject to Eells's type of objection, since it follows that $\Pr(T/T \vdash E) = \Pr(T)/\Pr(T \vdash E)$, which means that if $\Pr(T) = \Pr(T \vdash E)$, then $\Pr(T/T \vdash E) = 1$. But the result that $\Pr(T \& (T \vdash E)) = \Pr(T)$ says that the agent is certain that if $T$ is true then $T \vdash E$. Since $T \vdash E$ is (probabilistically) a consequence of $T$, $\Pr(T) \leqslant \Pr(T \vdash E)$, with strict inequality typically holding. Thus, the analogue of Eells's objection for the present derivation lacks bite.

Alas, the problem of old evidence in the perihelion case remains unresolved. The meaty condition (A3) says that upon writing down his theory, Einstein was certain that it implied a definite result about the advance of the perihelion of Mercury. But the historical evidence goes against this supposition. Indeed, Einstein's published paper on the perihelion anomaly contained an incomplete explanation, since, as he himself noted, he had no proof that the solution of the field equations he used to calculate the perihelion was the unique solution for the relevant set of boundary conditions.

Assume that Einstein's degree of belief in GTR, conditional on the theory's giving the correct prediction for the perihelion of Mercury, was greater than his degree of belief in the theory, conditional on the theory's giving no definite prediction about the perihelion. If we let $T \vdash N$ stand for $\neg(T \vdash E) \& \neg(T \vdash \neg E)$, the assumption amounts to replacing (A3) with

$$\Pr(T/T \vdash E) > \Pr(T/T \vdash N). \tag{A3'}$$

Then (A1), (A2), (A3'), and (A4) together imply that $\Pr(T/T \vdash E) > \Pr(T)$.

*Proof*

$$\Pr(T) = \Pr(T/T \vdash E) \times \Pr(T \vdash E) + \Pr(T/T \vdash \neg E) \times \Pr(T \vdash \neg E)$$

$$+ \Pr(T/T \vdash N) \times \Pr(T \vdash N).$$

By (A1.a) and (A4), the second term on the right-hand side is 0. And since $\Pr(T \vdash E) + \Pr(T \vdash N) \leqslant 1$, $\Pr(T \vdash E) < 1$ (by (A2)), and $\Pr(T/T \vdash E) > \Pr(T/T \vdash N)$ (by (A3')), the equality cannot hold if $\Pr(T) \geqslant \Pr(T/T \vdash E)$.

While the doubt left by Eells's analysis may not have been completely resolved in favor of the GJN approach, I think enough has been said to make it plausible that in an interesting range of cases, learning $T \vdash E$ can serve to boost confidence in $T$.

## 5   The Inadequacy of the Garber, Jeffrey, and Niiniluoto Solution

For those Bayesians who have been persuaded by Garber, Jeffrey, and Niiniluoto of the need to humanize their doctrine, the way is now open to search through the Einstein archives for evidence that in November of 1915 Einstein's beliefs conformed to (J1) through (J4) or to one of the alternative schemes (A1) through (A4) or (A1), (A2), (A3'), (A4). Suppose that the findings are positive (alternatively, negative). Would the problem of old evidence with respect to GTR and the perihelion of Mercury have thereby been shown to have a positive (alternatively, negative) solution? Not at all. The original question was whether the *astronomical data E* confirmed GTR (for Einstein, if you like). Garber, Jeffrey, and Niiniluoto replace this question with the question of whether Einstein's learning that $T \vdash E$ raised his confidence in the theory. Not only are the two questions not semantically equivalent; they are not even extensionally equivalent. We can say without a shadow of a doubt that for Einstein $E$ did confirm $T$. But we have to be prepared for the archival finding that the conditions needed to prove that $\Pr(T/T \vdash E) > \Pr(T)$ fail for him.

The point becomes clearer when we shift from Einstein to others. We now want to say that the perihelion phenomenon was and is good evidence for Einstein's theory. But along with most students of general relativity, the first thing we may have learned about the theory, even before hearing any details of the theory itself, was that it explains the perihelion advance. So there never was a time for us when $\Pr(T \vdash E) < 1$.

Moreover, even if the two questions

Does $E$ confirm $T$ for person $P$?

Does learning $T \vdash E$ increase $P$'s degree of belief in $T$?

should stand or fall together, there is no guarantee that the strength of confirmation afforded by $E$ is accurately measured by the boost given to degrees of belief by learning $T \vdash E$. This matter is connected to the issue of whether $E$ can confirm a theory designed to explain $E$.[13] It will be helpful here to distinguish three senses in which person $P$ might be said to have designed $T$ so as to explain $E$.

1. When $P$ created $T$, he was motivated by a desire to explain $E$.
2. Before settling on $T$, $P$ examined and rejected alternative theories that failed to explain $E$.
3. In arriving at $T$, $P$ went through an explicit chain of reasoning that started with $E$ and worked back to $T$.

As we move from (1) to (3), it becomes less and less surprising to $P$ that $T \vdash E$, and therefore $P$'s learning that $T \vdash E$ gives a smaller and smaller boost to his degree of belief in $T$. We have already seen that Einstein satisfied (1). That he satisfied (2) is indicated by the fact that he wrote to Sommerfeld in November of 1915 that one of the reasons he abandoned a previous theory, constructed with the help of his friend Marcel Grossmann, was that it yielded an advance of only 18" per century for Mercury's perihelion.[14] But this piece of personal history does not seem to have diminished the confirmational value of $E$, as opposed to $T \vdash E$, for Einstein. Nor would the discovery that Einstein also satisfied (3) show that $E$ had no confirmational value for Einstein or his fellow scientists.[15]

Substituting a tractable problem for an intractable one is a time-honored tactic. The tactic is fruitful if the solution to the tractable problem illuminates the original problem. In this case, however, the solution given by GJN to the Bayesian learning problem for humanized agents fails to speak to the original problem. Further, the so-far intractable part of the problem of old evidence is just as much a problem of new theories as of old evidence. How probability is to be assigned to the newly minted theory is a question that must be answered before we can begin to worry about whether and how the probability of $T$ is boosted by $E$, by $T \vdash E$, or by whatever. The problem of new theories will be touched upon in section 6

and discussed in more detail in chapter 8. But before I close this section, it will be helpful to sketch a non-Bayesian account of why the perihelion data does constitute good confirmation of GTR.

Such an account could appeal to at least four facts: (1) that GTR yields the exact value of the anomalous advance, (2) that it does so without the help of any adjustable parameters, (3) that the perihelion phenomenon provides a good bootstrap test of GTR, and (4) that dozens of attempts were made within both classical and special relativistic physics to resolve the anomaly, all of which failed.[16] By way of explaining (3), assume that the exterior field of the sun is stationary and spherically symmetric. Then the line element can be written as

$$ds^2 = [1 - (\alpha m/r) + (2\beta m^2/r^2) + \cdots]dt^2$$
$$- [1 + (2\gamma m/r) + \cdots](dx^2 + dy^2 + dz^2),$$

where $m$ is the mass of the sun, $r^2 = x^2 + y^2 + z^2$, and $\alpha$, $\beta$, $\gamma$ are undetermined parameters. Einstein's GTR requires that $\alpha = \beta = \gamma = 1$. The first-order red shift depends only on $\alpha$, while the bending of light depends only on $\alpha$ and $\gamma$. By contrast, the advance of the perihelion of a planet depends upon all three parameters, and so the perihelion data helps to pin down a parameter left undetermined by the other two classical tests.

The point here is that without doing any Bayesian calculations and without solving the Bayesian problem of old evidence, we can recognize on independent grounds the confirmational virtues of the perihelion data. Of course, there is nothing to block Bayesians from taking into account the factors enumerated above. But how these factors can be made part of Bayesian calculations in the context of old evidence remains to be seen.

## 6   New Theories and Doubly Counting Evidence

Despite my rejection of the GJN substitution move, I agree with the main thrust of their humanized Bayesianism: namely, a realistic theory of confirmation must take into account nonempirical learning. But while Garber, Jeffrey, and Niiniluoto address the learning of logicomathematical facts, they, like most of Bayesian authors, are silent about the learning of new theories, despite the obvious importance of such learning for an understanding of scientific change.[17] Indeed, I would venture that the problem of new theories presents both a more interesting challenge and a more

interesting opportunity for Bayesians than does the original problem of old evidence.

In Bayesian terms, the introduction of new theories causes a humanized Bayesian agent (who fails (LO2)) to shift from a probability function Pr, operative before the introduction, to a new function Pr', operative after the introduction. And typically, Pr' is not derived from Pr by any straightforward conditionalization process. How this transition is or ought to be managed is a matter that I will take up in chapter 8. For present purposes, the details of how Pr' is generated are irrelevant.

The problem of new theories presents the opportunity to further explore the slogan that a theory $T$ is not confirmed by evidence $E$ that $T$ was designed to explain. Suppose that $E$, whether fresh or stale, leads to the proposal of a new theory $T$, and suppose that this new $T$ is assigned a nonzero probability relative to the new Pr' function. Then since the Pr' assignments were made in light of $E$, it would seem to be double counting the evidence to take it to confirm $T$ in the sense of raising the Pr' probability of $T$.[18] That is the kernel of truth in the slogan. (Of course, if the agent fails logical omniscience (LO1) his assignment Pr'($T$) may not accurately reflect the evidential import of $E$, for he may fail to know that $T \vdash E$, and upon learning the implication, he may change his degree of belief in $T$ à la Garber and Jeffrey. But this does not undercut the prohibition against doubly counting evidence.)

It is worth noting that, looked at from the perspective of new theories, the problem of old evidence is not a problem at all but merely an application of the methodological truism that evidence should not be doubly counted. But looked at from the ex post facto perspective, the problem of old evidence is a real problem, since we want to affirm that $E$ does after all confirm or support $T$.

## 7   Conclusion: A Pessimistic Resolution of the Old-Evidence Problem

The recognition that the interesting residual problem of old evidence arises from the problem of new theories is important, for it automatically undercuts some of the proposed treatments of old evidence. Suppose that the problem of a new theory has been resolved in that in reaction to the introduction of $T$ the Bayesian agent chooses, in some appropriate way, a new probability function Pr' such that Pr'($T$) > 0. In this setting, we

cannot follow Howson's (1984, 1985) prescription for resolving the old-evidence problem by computing the difference between what the agent's degree of belief in $T$ would have been if his total knowledge at the time $T$ was introduced had been $K - \{E\}$, and what his degree of belief in $T$ would have been were he subsequently to come to learn $E$.[19] At least we cannot take this computation to be given by the comparison of $\mathrm{Pr}'(T/K)$ and $\mathrm{Pr}'(T/K - \{E\})$, since both of these probabilities are equal to $\mathrm{Pr}'(T)$, unless the very introduction of $T$ caused the agent to become uncertain about what he previously regarded as certain and accordingly to assign $\mathrm{Pr}'(K) < 1$. Howson's prescription is relevant for Bayesian agents who are logically omniscient in sense (LO2) and who change their belief functions only by conditionalization. But it is not a prescription that will cure the tough version of the old-evidence problem for agents who fail (LO2) and who resort to non-Bayesian shifts in their belief functions when new theories are introduced.

There seem to me to be only two ways to deal with the residual problem of old evidence. The first is to ignore it in favor of some other problem. That, in effect, is the route pioneered in the GJN approach. A perhaps better motivation for going this route derives from the view that ultimately our goal in scientific enquiry is to choose among competing theories, and for that choice, what matters are the relative values of the probabilities of the theories conditional on the total evidence, old as well as new (see, however, chapter 7).

But if the problem is not to be ignored, the only way to come to grips with it is to go to the source of the problem: the failure of (LO2). There are in turn two strategies for coping with the failure of (LO2), both involving counterfactual degrees of belief based not just on counterfactual evidence sets but on counterfactual probability functions. One version, already mentioned in section 2 above, imagines that the agent is empirically deficient as well as logically deficient. It imagines that the agent didn't know $E$ and asks what, in these circumstances, the agent's degree of belief in $T$ would have been when $T$ was introduced, and then it compares that number with what the agent's subsequent degree of belief in $T$ would have been had he then learned $E$. The computation is thus done using not the probability function $\mathrm{Pr}'$ actually adopted by the agent upon the introduction of $T$ but a hypothetical function. The other version imagines what the agent's degree of belief in $T$ would have been *ab initio* if he were not logically deficient but were a superhuman calculator satisfying (LO1) and

(LO2), and then it compares this number with the degree of belief this supercalculator assigns after learning $E$. This calculation involves a hyper-hypothetical probability function. I have no doubt that counterfactual enthusiasts can concoct ways to get numbers out of one or both of these scenarios. But what has to be demonstrated before the problem of old evidence is solved is that the numbers match our firm and shared judgments of the confirmational values of various pieces of old evidence. It would be quite surprising if such a demonstration could be given, since the counterfactual probabilities and thus the counterfactual incremental boosts in confirmation will vary greatly from one person to another. Hope springs eternal, but even if the hope is realized the Bayesian account of confirmation retains a black eye for being forced to adopt such a complicated and dubious means of accommodating such a simple and common phenomenon in scientific inference.

For results using concepts of partial exchangeability, see Diaconis and Freedman 1980. For results using ergodicity, see von Plato (1982).

2. For simplicity I have suppressed reference to the background knowledge.

3. But see section 7 for a discussion of just how weak this notion of instance induction is.

4. For another response to Gillies's form of the Popper and Miller argument, see Dunn and Hellman 1986. Yet another move the inductivist can make is to use a different measure of inductive support (see Redhead 1985). Popper and Miller's responses to criticism are to be found in their 1987 paper. See also Howson 1990b.

5. This apt phrase of 'content cutting' is due to Ken Gemes.

6. But see chapter 6, where cases of the nonobjectivity of likelihoods are discussed.

7. Here I am indebted to Philip Kitcher. However, we differ on what philosophy of science can be expected to provide by way of a solution (see chapter 8).

8. The definitions of 'grue' given in the literature are confusing. Goodman defines 'grue' so that "it applies to all things examined before $t$ just in case they are green but to other things just in case they are blue" (1983, p. 74). In the preface to the fourth edition of *Fact, Fiction, and Forecast* Hilary Putnam says that an object is grue "if it is either observed before a certain date and is green, or it is not observed before that date and is blue" (Goodman 1983, p. vii). To get the effect of a shift in color after the magic date that most readers expect from grue, we can employ two-place predicates: $Ext$, meaning that $x$ is an emerald at $t$, and $Gxt$, meaning that $x$ is grue at $t$, that is, that either $t$ is no later than 2000 and $x$ is green or else $t$ is later than 2000 and $x$ is blue. The hypothesis that all emeralds are grue then reads, $(\forall x)(\forall t)(Ext \to Gxt)$. Alternatively, we can consider a temporally ordered series of objects $a_1$, $a_2, \ldots$ and take the grue hypothesis to be $(\forall i)(Ea_i \to Ga_i)$, where $Ga_i$ means that $i \leqslant 2000$ and $a_i$ is green or else $i > 2000$ and $a_i$ is blue. The latter alternative is followed here.

9. Goodman might be read as making a stronger demand on genuine confirmation, namely, that it be homogeneous. One way to unpack the notion of homogeneity is through Hempel's special consequence condition: if $E$ confirms $H$, then $E$ confirms any logical consequence of $H$. But as argued in chapter 3, this condition is in general unacceptable if confirmation is interpreted in terms of incremental boosts in probability. A more attractive explication of homogeneity is in terms of exchangeability of the instances of $H$. But to see Goodmanization whenever there is a failure of homogeneity in this sense is to see Goodmanization throughout science (see the discussion below for more about the connection between exchangeability and Goodman's problem). Ken Gemes (1990b) defines a notion of "real confirmation" requiring that for $E$ to really confirm $H$, it must incrementally confirm each "content part" of $H$. Gemes attempts to provide a syntactical definition of the notion of content part that does justice to the intuition that not every logical consequence of $H$ is a genuine content part. To leave aside technical problems with Gemes's definition, my main concern is that his notion of real confirmation is too demanding to be of much use in real scientific examples. For instance, none of the classical tests of Einstein's general theory of relativity or any of the more recent tests I am aware of can plausibly be said to confirm every content part of the theory, and this even though these tests are commonly held to provide good confirmation of the theory. Of course, one could say that individually or collectively these pieces of evidence do confirm every content subpart of some significant part of Einstein's theory. But by the same token, green emeralds confirm every subcontent part of some significant part of 'All emeralds are grue'. In sum, I am unconvinced that there is any workable and useful notion of homogeneous confirmation to be had. And I am convinced that the tools already developed are adequate to capture the valid lessons to be drawn from Goodman's examples.

10. The reason for the qualification 'future-moving' will become evident below.

11. A few years later Carnap abandoned the notion that there is a unique correct confirmation function (e.g., $c^*$), and in his later years he tended more and more toward personalism.

However, he remained a tempered personalist, for he thought that inductive common sense supplied constraints over and above those of the probability calculus (see Carnap 1980). The problem, of course, is that Carnap's inductive common sense did not agree with the inductive common sense of others. Did he think that there is some overriding *inductive common sense* to which, by some special faculty, he had access? I discuss below how a Carnap aligned with Bayesian personalism should have responded to Goodman's treatment of induction.

12. One of the earliest expressions of this point is to be found in Teller 1969.

13. The word 'grue' was introduced only at a later date.

14. Carnap's papers are part of the Archives for Scientific Philosophy, University of Pittsburgh. The quotation is from document no. 084-19-34, dated January 27, 1946. Quoted by permission of the University of Pittsburgh and C. G. Hempel. All rights reserved.

15. Carnap refereed the manuscript of *Fact, Fiction, and Forecast* for Harvard University Press. He recommended it for publication. Carnap's letter to the press is preserved as document no. 084-19-02 in the Archives for Scientific Philosophy, University of Pittsburgh.

## Chapter 5

1. The background knowledge $K$ is suppressed here for the sake of simplicity.

2. The present account skates over some of the historical details. For a full account, see Earman and Glymour 1991.

3. Temporal subscripts on the Pr function will be dropped whenever no confusion will result.

4. The doubts are discussed in detail in chapter 6 below.

5. $\Pr(T \ \& \ \neg E/\neg E) = \Pr(\neg E/\neg E) \times \Pr(T/\neg E) = \Pr(T/\neg E)$ (by (CP3) and (CP2) of appendix 1 of chapter 2). Thus $\Pr(T/\neg E) = 0$ when $T \models E$. When in addition $\Pr(E) = 1$, $\Pr(T/E) = \Pr(T \ \& \ E)/\Pr(E) = \Pr(T)$. So under the stated conditions, $\hat{C}(T, E) = \Pr(T)$.

6. See van Fraassen (1988) who mentions this line without advocating it.

7. See Earman and Glymour 1991 for details and references.

8. Similar problems arise if confirmatory power is measured in other ways, e.g., by Gaifman's (1985) $(1 - \Pr(T))/(1 - \Pr(T/E))$.

9. Here I am using the terminology of Eells 1985. The approach sketched here can also be found in Skyrms 1983.

10. That this was a genuine piece of learning for Einstein is indicated by the fact that he spent about a week of hectic calculation to derive the prediction of perihelion advance from GTR (see Earman and Glymour 1991). And when he found his prediction matched the observed anomaly, he suffered heart palpitations (see Pais 1982, p. 253).

11. Horwich (1982) and Howson (1984, 1985) both advocate that for old evidence, incremental support should be measured in terms of counterfactual degrees of belief. It is not clear, however, at which version of the old-evidence problem their constructions are aimed. If they are aimed at the problem of *old new evidence*—i.e., $T$ was formulated before the discovery of $E$ but it is now later and $\Pr(E) = 1$—then the constructions amount to much the same thing as recommended by Skyrms (1983) and Eells (1985). But if the constructions are aimed at the problem of *new old evidence*—i.e., $E$ was known before the formulation of $T$ and it is now the time of the formulation of $T$ (or barely after it)—or the problem of *old old evidence*—i.e., $E$ was known before the formulation of $T$ and it is now some time subsequent to the formulation—then they are subject to the difficulties discussed in the present section and also in section 7 below.

12. For present purposes, nested iterations of '⊢' will be ignored.

13. For various opinions on this matter, see Howson 1984 and Worrall 1978, 1989. See also the discussion in section 8 of chapter 4.

14. See Hermann 1968, p. 32.

15. See R. Miller 1987, chap. 7, for similar sentiments.

16. For details, see Roseveare 1982 and Earman and Glymour 1991.

17. The problem posed by new theories for orthodox Bayesianism is touched upon by Teller (1975, pp. 173–174). It is discussed explicitly by Chihara (1987). One of the few references in the statistics literature I know of is Leamer 1978, where the problem is discussed under the rubric of "data instigated" hypotheses.

18. See Leamer 1978 for a discussion of the prohibition against doubly counting evidence.

19. Some further remarks about Howson's $K - \{E\}$ are in order. Sometimes he seems to think of $K$ as a discrete set of elements from which $E$ can be plucked. But normally $K$ is treated as being closed under logical implication. (If (LO1) fails, this is an unrealistic assumption, of course.) In this case we might try to get at the relevant sense of $K - \{E\}$ along the following lines. Call $K'$ an intermediary between $K$ and $E$ just in case $K'$ is stated in the vocabulary of $K$, $K \models K'$, and $K' \not\models E$. If there is a unique strongest intermediary, take this to be $K - \{E\}$. There seems to be no guarantee, however, that there will be a unique strongest intermediary. See also Chihara 1987, pp. 553–554.

## Chapter 6

1. Here I am indebted to an unpublished lecture of Peter Railton's, delivered at the Pittsburgh Center for the Philosophy of Science in spring 1990.

2. On this matter, see the discussion in chapters 5 and 8.

3. See Jaynes 1957, 1968. For a critical assessment of the maximum entropy principle, see Seidenfeld 1979, 1986.

4. This is a point that has been repeatedly stressed by Wes Salmon. The Bayesian must deny that these plausibility arguments represent a logically prior form of reasoning that delivers plausibility assessments (see chapter 2).

5. Moving to a rule that determines Pr values not just on a partition but on the entire probability space would be a big step toward apriorism.

6. I have borrowed this example from Wes Salmon.

7. For a discussion of this form of tempered personalism, see Shimony 1970. Note, however, that the sort of criticism given in chapter 8 of Lehrer and Wagner's (1981) proposal for a rule that forms a consensus by means of a weighted aggregation of opinions also applies to this form of tempered personalism.

8. This move falls in with the definitional solution mentioned below in section 9 and discussed critically in chapter 8.

9. As explained in appendix 2 to chapter 2, the strong law of large numbers relies on the countable additivity of the measure on the space of infinite sequences of coin flips. But countable additivity of the degree of belief function Pr is not assumed here. If one wanted to eschew countable additivity altogether, one could fall back on the form of the weak law of large numbers that says that as the number of flips goes to infinity, the probability that the observed frequency of heads differs from the true chance value by any $\varepsilon > 0$ goes to zero. The reader is invited to try to apply this fact to derive a conclusion about merger of opinion.

10. A *random variable* is a map $X : \Omega \to \mathbb{R}$ such that for every Borel set $A \subseteq \mathbb{R}$, $\{\omega \in \Omega : x(\omega) \in A\}$ is in $\mathscr{F}$. Recall that a *field* is a collection of subsets of $\Omega$ closed under complementation and finite unions, and that a $\sigma$ field is closed under countable unions. See appendix 2 of chapter 2 for more details.

11. $E(X)$ is the expectation value of the rv $X$. $E(X/\mathscr{F}_n)$ stands for the conditional expectation value of $X$ on $\mathscr{F}_n$. If $\mathscr{F}_n$ is the $\sigma$ field generated by the partition $\{B_j\}$ of $\Omega$, then $E(X/\mathscr{F}_n)$ is defined as the rv such that $E(X/B_j)$ has constant value on each set of the partition.

12. For a different derivation of results similar to those in the theorem, see Schervish and Seidenfeld 1990.

13. The difference between the strong and the weak forms of merger of opinion lies in the order of the quantifiers. For an evidence matrix $\Phi = \{\varphi_i\}$, define the conditional distance measure on probability functions as

$$\mathrm{dist}_{\Phi,n,w}(\mathrm{Pr} - \mathrm{Pr}') \equiv \sup_\psi |\mathrm{Pr}(\psi / \underset{i \leqslant n}{\&} \varphi_i^w) - \mathrm{Pr}'(\psi / \underset{i \leqslant n}{\&} \varphi_i^w)|.$$

Then part (2) of the theorem proves weak merger: for any separating $\Phi$, for a.e. $w \in \mathrm{Mod}_{\mathscr{L}}$, any $\varepsilon > 0$, and for any equally dogmatic Pr and Pr', there exists an $N$ such that for any $m \geqslant N$, $\mathrm{dist}_{\Phi,m,n}(\mathrm{Pr} - \mathrm{Pr}') < \varepsilon$. Strong merger requires that … there exists an $N$ such that for any equally dogmatic Pr and Pr'….

14. See, however, the discussion below in section 7. An *atomic observation sentence* is a formula of the form $P i_1 i_2 \dots i_k$ or $f(i_1, i_2, \dots, i_k)$ where '$P$' and '$f$' are respectively a $k$-ary observation predicate and a $k$-ary function symbol and the $i$'s are numerals.

15. It is conceivable that the opinions of Bayesian agents regarding a hypothesis could merge in the sense that the absolute difference of the posterior probabilities of any two agents goes to 0 without there being a convergence to certainty on the hypothesis. This raises the question of what conditions need to be assumed in order to prove that merger implies convergence to certainty. As a simple illustration, suppose that $H_1$ and $H_2$ are treated as mutually exclusive and exhaustive. And suppose that the class of probability functions $\{\mathrm{Pr}\}$ is *rich enough* that it contains $\mathrm{Pr}_1$ and $\mathrm{Pr}_2$, which differ on the prior odds but agree on the posterior odds for any evidence from the evidence matrix $\Phi = \{\varphi_i\}$. That is,

$$c_1 \equiv \mathrm{Pr}_1(H_1)/\mathrm{Pr}_1(H_2) \neq \mathrm{Pr}_2(H_1)/\mathrm{Pr}_2(H_2) \equiv c_2$$

but

$$R_n \equiv \mathrm{Pr}_1\left(\underset{i \leqslant n}{\&} \pm \varphi_i / H_1\right)\Big/\mathrm{Pr}_1\left(\underset{i \leqslant n}{\&} \pm \varphi_i / H_2\right) = \mathrm{Pr}_2\left(\underset{i \leqslant n}{\&} \pm \varphi_i / H_1\right)\Big/\mathrm{Pr}_2\left(\underset{i \leqslant n}{\&} \pm \varphi_i / H_2\right)$$

for all $n$. Now suppose that merger of posterior opinion takes place in the sense that there is an $r$ such that $\mathrm{Pr}_{1,2}(H_1 / \underset{i \leqslant n}{\&} \pm \varphi_i) \to r$ as $n \to \infty$. (Note that this is stronger than requiring merger, in the sense that the absolute value of the difference of the posterior probabilities goes to 0, which would allow the opinions to oscillate without settling down to a definite value.) Then $r$ must be 1 or 0. For

$$\mathrm{Pr}_1\left(H_1 / \underset{i \leqslant n}{\&} \pm \varphi_i\right) - \mathrm{Pr}_2\left(H_1 / \underset{i \leqslant n}{\&} \pm \varphi_i\right) = (c_1 - c_2)/(c_1 + c_2 + R_n + c_1 c_2 / R_n).$$

Since this quantity goes to 0 as the limit $r$ is approached, $R_n$ must go either to 0 or to $\infty$, which imply respectively that the posterior probability of $H_1$ goes to 0 or 1. For details and a generalization to a countable number of hypotheses, see Hawthorne 1988. Under what other conditions does merger of opinion imply convergence to certainty?

16. I implicitly assumed in section 2 and will explicitly assume in this section that the observational/theoretical distinction can be drawn linguistically by means of a bifurcation of the empirical vocabulary of the language into observational and theoretical terms, the former of which denote directly observable properties or relations. An observation sentence is then