Putting on the Garber Style? Better Not

Colin Howson*†

This article argues that not only are there serious internal difficulties with both Garber's and later 'Garber-style' solutions of the old-evidence problem, including a recent proposal of Hartmann and Fitelson, but Garber-style approaches in general cannot solve the problem. It also follows the earlier lead of Rosenkrantz in pointing out that, despite the appearance to the contrary which inspired Garber's nonclassical development of the Bayesian theory, there is a straightforward, classically Bayesian, solution.

1. The 'Old Evidence Problem'. The 'old evidence problem' is reckoned to be a problem for Bayesian analyses of confirmation in which evidence E confirms hypothesis H just in case $P(H|E) > P(H)$. It is reckoned to be a problem because in such classic examples as the rate of advance of Mercury's perihelion (M) supposedly confirming general relativity (GR), the evidence had been known before the theory was proposed; thus, before GR was developed P(M) was and remained equal to 1, and Bayes's Theorem tells us that therefore $P(GR|M) = P(GR)$. The failure is all the more embarrassing since M was not used by Einstein in constructing his theory (he wrote that he had heart palpitations when he discovered that M followed straightforwardly from GR) and so to all intents and purposes was 'new' evidence vis-à-vis GR.

Among notable attempts to solve the problem was Garber's, building on the suggestion of Glymour (1980) that what confirmed GR was not M itself but that GR entails M, a fact discovered of course by Einstein himself. In developing this idea into a general solution, Garber expanded the domain of Pr to include not only H and E for arbitrary H and E but also '$H \vdash E$' as an additional proposition capable of taking probability less than 1 (while $P(E) = 1$) and sought to show that by imposing appropriate 'logical' constraints on the probabilistic behavior of '$H \vdash E$' he could prove that $P(H|\text{`}H \vdash E\text{'}) > P(H)$.

Received May 2016; revised November 2016.

*To contact the author, please write to: 16 Tranby Avenue, Toronto, ON M5R 1N5, Canada; e-mail: howson.colin@gmail.com.

†I would like to thank two anonymous reviewers for their very helpful advice.

Philosophy of Science, 84 (October 2017) pp. 659–676. 0031-8248/2017/8404-0003$10.00
Copyright 2017 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/693466 Published online by Cambridge University Press

In this he was only partly successful, but a later refinement due to Earman (1992) yielded the required result.

A recent paper by Hartmann and Fitelson (2015) takes further the Garber-style strategy of expanding the domain of the probability function to include suitable metalevel propositions and seeks to show that it succeeds in a broader environment where P(E) is not assumed to be 1 or indeed to take any particular value.[1] In what follows I argue that there are severe technical and explanatory difficulties with Garber-style approaches in general, including Hartmann and Fitelson's, and that in any case no Garber-style approach is in principle capable of solving the old evidence problem. In the final section I show that a minimalist version of Objective Bayesianism does straightforwardly solve the problem, and I argue that only it is capable of furnishing a satisfactory probabilistic theory of inductive inference.

2. What's Wrong with Garber-Style Approaches? A prima facie problematic feature of Garber's approach is that, as he himself pointed out, it appears to require a nonclassical probability function to assign values, and in particular a value less than 1, to statements of the form '$H \vdash E$' when they are true. Garber partly circumvented this by taking '$H \vdash E$' to be an atomic sentence in a propositional language L closed off under the Boolean operations, on which a wholly classical probability function is defined. However, as Eells pointed out, there is then the anomalous situation that there may be sentences U and V of L such that '$U \rightarrow V$' is a tautology and hence has probability 1, while there are two other sentences H and E of exactly the same respective truth functional for such that '$H \vdash E$' is atomic in L and has probability less than 1 (1990, 214, 215). Thus, it may be that the probability distribution over the items in L is coherent but only when the bets are not on the content of those expressions: it is a case of *selective blindness*. Clearly, a similar situation can arise if the language in question is first order.

There is another curious feature of the 'Garber-style' claim that it is some logico-epistemological fact, rather than an empirical one, that confirms a scientific theory H—in Garber's own treatment the logico-epistemological fact in question is of course just a logical one, namely, that $H \vdash E$. Now whether the language of H and E is propositional or first order, the sentence '$H \vdash E$' is actually a disguised *pure-mathematical* assertion, indeed essentially an arithmetical one. In the first-order case, under a Gödel numbering mapping the formal structure of a first-order language into the standard model of first-order arithmetic, Gödel's famous incompleteness proof shows that the entail-

1. They introduce two binary variables X and Y, where X is "the hypothesis H is a satisfactory explanation of the evidence E" and Y is "the nearest competitor to H is a satisfactory explanation of E," and show that if X and Y satisfy four allegedly plausible constraints, then $P(H|X) > P(H)$ (Hartmann and Fitelson 2015).

https://doi.org/10.1086/693466 Published online by Cambridge University Press

ment relation is a $\sum_1$ arithmetical relation (it is of the form $\exists xA(x,y,z)$, where $A(x,y,z)$ is recursive); in the propositional case it is recursive. Thus, what Garber, following Glymour, is in effect claiming is that a scientific theory may be confirmed, Bayesianwise, by a statement that is true for purely arithmetical reasons. That sounds bizarre, and indeed to my mind it is.

In Hartmann and Fitelson's treatment, also, incoherence lurks not far beneath, if not above, the surface. A probability distribution over a propositional language in which H, X, and Y are the only atomic sentences might be coherent, but not when, so to speak, we look inside those atoms to see what they assert. Thus, given the sorts of constraints determining what is and what is not a satisfactory explanation that we would ordinarily assume incorporated in background information and so 'built into' P, it should be an entailment from that information that H is a satisfactory explanation of E if in fact it is. In that case, if X is true then, modulo that background information, it is equivalent to a logical truth T. But since by assumption (to solve the initial problem) $P(X) < 1$, we must by coherence have $P(T) < 1$, which is incoherent. Hartmann and Fitelson say that they "suspect" that their "approach can be made compatible *even* with subjective Bayesian accounts of scientific explanation" (2015, 717 n. 5). In view of the above, that is rather more than doubtful.

Not only are there substantial internal problems with Garber-style approaches, but there is a fundamental, and to my mind devastating, explanatory objection. Ever since Einstein published his proof that GR (which I take to be simply the assumption of a static spherically symmetric field around the sun with Mercury a material point traveling along a geodesic) yielded a straightforward explanation of the anomalous amount of Mercury's perihelion precession "in full agreement" with the astronomical data (Einstein 1915, 839), it has been almost universally assumed that GR was strongly confirmed by the observed perihelion motion, including by Einstein himself: "I found an important confirmation of that radical Relativity theory; *it exhibits itself namely in the secular turning of Mercury in the course of its orbital motion, as was discovered by Le Verrier*" (831; my emphasis). Indeed, of the three early tests of GR, gravitational red shift, deflection of light by the sun, and Mercury's perihelion motion, the last is regarded by (as far as I am aware) practically the entire physics community then and now as the most powerful in its confirming power. Earman sums up the general view: "we want to say that the perihelion phenomenon did (and does) lend strong support to Einstein's theory" (1992, 121).

Arguably a desideratum for an adequate confirmation theory is that when there is such unanimity among the principal actors, it should be modeled within that theory, or if not, some convincing reason should be supplied why they are mistaken. Taking his lead from the subjective Bayesian theory, Garber does, implicitly, argue that they were mistaken, and they were mis-

https://doi.org/10.1086/693466 Published online by Cambridge University Press

taken because they failed to identify the true reason for their increase in confidence in GR: "If old evidence can be used to *raise* the probability of a new hypothesis, then it must be by way of the *discovery* of previously unknown logical relations. In the cases that give rise to the problem of old evidence, we are thus dealing with circumstances in which hypotheses are confirmed not by the empirical evidence itself, but by the discovery . . . that $h \vdash e$" (1983, 120).[2]

But this is a non sequitur: while it is almost certainly true that the immediate cause of the increase in confidence in GR was the discovery that GR predicted the observations of Mercury's deviant perihelion motion (M), it does not follow that M itself did not confirm GR. As Rosenkrantz pointed out in a paper published in the same year as Garber's, according to a Bayesian confirmation theory of long pedigree—going back to Bayes himself—M provably confirms GR (and to a very large degree) in the usual Bayesian sense of raising its probability above the prior (assumed nonzero). We will see in due course that a simple Bayes's Theorem calculation shows that M strongly confirmed GR because M was strongly expected given GR (because it was entailed by GR) and extremely unlikely given the only theory rival to GR. It follows that, without any appeal to Garber-style modifications of the formalism, that calculation will explain both how the discovery that GR entailed M justified an augmented confidence in GR and why it was nevertheless M itself that did the confirming.[3] In this light, the failure of Garber's own theory to account for the virtually unanimous opinion of the science community about the confirming nature of M—indeed, implicitly denying it—seems nothing less than a massive explanatory failure.

All this will of course depend on showing that '$P(M) = 1$' can be denied within the sort of Bayesian theory Rosenkrantz was advocating without sacrificing either consistency or the intuitive plausibility of Bayesian reasoning famously summed up by Laplace as "common sense reduced to a calculus"

2. To have '$h \vdash e$' confirming h is impossible unless '$h \vdash e$' has a probability less than 1. But in standard subjective Bayesianism a logico-mathematical proposition has probability 1 if true, whence Garber famously—and (I will argue) wrongly—diagnosed the old evidence problem as due to Bayesianism's presumption of "logical omniscience" (1983, 106). Good prepared the ground for the charge of logical omniscience when he claimed that learning that H implies E should be represented by a transition from $P(E|H) < 1$ to $P(E|H) = 1$ within a theory of what he called "evolving," or "dynamic," probability (107; the paper referred to was first published in 1977).

3. Nothing I have said should be taken to imply that this was Einstein's own reasoning, which almost certainly it was not. All that it purports to show is how a fully equipped Bayesian robot, of a sort whose acquaintance we will make later, might reason when informed that GR entails M. It does not follow either that because the discovery caused an increased degree of belief in GR the discovery itself confirmed GR. A drug might also cause an increased belief in GR.

https://doi.org/10.1086/693466 Published online by Cambridge University Press

(*le bon sens réduit au calcul*). To that end, it will be helpful to make a closer acquaintance with that theory.

3. A Tale of Two Bayesianisms. Now usually referred to as Objective Bayesianism, it was *the* Bayesian theory until subjective Bayesianism gained a wider audience after the Second World War, with a roll call of distinguished advocates who were largely working scientists united in their view that the rules of probability furnish the logic of scientific inference. And the reference to logic is not idle: according to a view promulgated vigorously by Jeffreys and Jaynes,[4] whose earlier inspiration was Keynes, this is a theory in which the rules of probability are interpreted, formally at any rate, as a generalization of deductive entailment, with a two-place function P(A|B) from pairs of propositions (A,B) with B consistent, into [0,1], and taking the value 1 if B entails A and 0 if B entails the negation of A. Those values are interpreted as degrees of plausibility of A given B, unlike the binary valid/invalid verdicts of classical deductive logic depending on logical structure alone. Jaynes himself described the theory as the operating system of what he called a "thinking computer," or "robot," incorporating rules of plausible reasoning in addition to those of deductive logic and mathematics (2003, 4). Unsurprisingly, among those rules are those of probability itself. In a proof that Jaynes adopted, and slightly adapted, as the foundation of his own theory, another physicist, Richard T. Cox, had shown that from two very general constraints on any acceptable real-valued measure m(A|B) (my notation) of what he called the "rational expectation" of A given B, there is a rescaling of it into the closed unit interval satisfying the finitely additive conditional probability axioms (1946).[5] Cox also saw his axioms as part of what he called "a logic of inference and enquiry," specifically an inductive logic (1976, 1), but I should stress again that the 'logic' to which all these authors appealed is quite distinct from deductive logic, which since the remarkable advances in the discipline inaugurated by Frege's seminal work has tended to appropriate the title uniquely to itself (throughout the eighteenth and nineteenth centuries it was common for 'logic' to subsume two distinct subdisciplines, deductive logic, also called 'the logic of certainty', and the theory of uncertain inference; Keynes, Jeffreys, Cox, and Jaynes in effect continued that earlier tradition into the twentieth century).

4. Jaynes's posthumously published book (2003) has the title *Probability Theory: The Logic of Science*.

5. In fact there are three constraints if sentences are the domain of the measure, when equivalent sentences are assumed to take the same value. Cox's proof was charged with inconsistency by Halpern (1996), but the charge seems to have arisen from a misreading of Cox and was later retracted (Snow [2001] contains a detailed discussion). Terenin and Draper (2015) show that, by adding domain-extension and continuity axioms, what they call the Cox-Jaynes derivation can be extended to deliver a countably additive probability.

https://doi.org/10.1086/693466 Published online by Cambridge University Press

Subjective Bayesianism is of course also a theory of plausible reasoning and uncertain inference (and was also credited with logical status by its two celebrated pioneers Ramsey and de Finetti; I will come back to this point in sec. 5), but as far as the theory itself is concerned its probabilities—all of them, not just the priors—merely reflect the personal beliefs of the agent subject to the weak constraint of consistency ('coherence') imposed by the probability axioms. To justify its name, then, there presumably ought to be some authentic standard of objectivity satisfied by the Objective theory. Its evaluation of so-called likelihoods does seem satisfactorily objective, as we will see in due course, but the priors have always presented a problem, so much so that the subjective theory with its frank confession of fallibility in such matters, combined with theorems appearing to prove that with accumulating evidence the priors eventually get 'washed out', in the Bayesian vernacular, became the more attractive option for many Bayesians. The tendency of the objectivist theory to see objectivity in what are often called 'informationless' priors unfortunately mired it in paradox and controversy, from the notorious Principle of Indifference onward through more recent appeals to reference priors, invariant priors, entropic priors, and others.[6] It also left it vulnerable to the charge that always trying to impose prior neutrality is misconceived because it may mean throwing away background information capable of making relevant prior discriminations between competing hypotheses. Is it sensible, for example, to demand prior neutrality between what qualified opinion judges a serious explanatory hypothesis and one without any such pedigree, cooked up simply to 'explain' an experimental result (I present an extreme example in the next section)?

The last observation/question suggests that a more promising avenue to pursue is to identify plausible indicators of explanatory potential and let the priors reflect these. Familiar candidates have been non-adhocness, simplicity (whose ancient pedigree is signaled in the Latin *simplex sigillum veri* [simplicity is the seal of truth]), analogy, unification, correspondence (reproducing existing theory in the limit of some parameter or parameters), and perhaps other qualities specifically relevant to the discipline, for example, possessing suitable symmetries.[7] That is all very well, you might say, but (*a*) do not these criteria beg the question of their reliability? And (*b*) even were that question

6. Purely logical criteria fail by themselves to determine any probability measures, as Carnap found in the context of extremely simple formal languages barely stronger than those of propositional logic. Jeffreys and Jaynes pioneered the use of invariance, and there have been some interesting results: e.g., in his discussion of Bertrand's famous chord problem Jaynes (1973) showed that invariance under the three transformation groups—scale, rotations, and translations—determines one of the three solutions Bertrand had claimed equally valid. But Jaynes also acknowledged that the method does not always work. In addition, it generates so-called improper (divergent) priors for scale parameters.

7. This list overlaps with that of Salmon (1990), another Objective Bayesian.

https://doi.org/10.1086/693466 Published online by Cambridge University Press

answered, how without some very dubious hand-waving do you extract prior probabilities from that list? Question *a* is of course just the old inductive-skeptical question, with its own vast literature that I have neither world enough nor time to attempt to summarize other than referring the reader to Goodman's classic discussion (1946, chap. 3) where he argues, I now think convincingly, that even the canons of deductive logic derive their authority from reflection on accepted deductive practice and that inductive rules should be judged by the same standard (although this is not necessarily to endorse his views about entrenchment).

As to *b*, while it would indeed seem impossible to condense those items into a single number, it is far from clear that this is a weakness. If we want a result valid for a single piece of evidence, like M, for example, then in the light of the observations above specifying a single number for the prior would arguably be overprecise (an observation applying also to interval-valued priors where the intervals themselves have sharp endpoints; for this reason the usual name 'imprecise probabilities' is not wholly apt),[8] with the result failing the test of robustness. Prominent objective Bayesians have not always seen lack of precision a problem: Jeffreys, who promoted a theory of simplicity measured by fewness of independent adjustable parameters (the Standard Model would not do very well), saw it only as inducing a prior ordering, and given that in most inferential contexts a reliable qualitative comparison is all that is wanted, that can be perfectly adequate. But Jeffreys also went further and showed that with evidence of a suitable kind a robust conclusion can be obtained by combining a purely objective function of the evidence with a prior characterized only very qualitatively. In the next section we will see how the combining operation works and that M is just the kind of evidence to generate such a conclusion.

4. Bayes's Theorem, Odds and Likelihoods. A combining operation is performed by Bayes's Theorem in its odds form:

$$\text{Odds}(H|E\&C) = \text{Odds}(H|C)\text{LR}, \tag{1}$$

where Odds(H|C) are the prior odds, C is relevant background information, and LR is the so-called likelihood ratio (sometimes also called the Bayes factor) equal to P(E|H&C)/P(E| ~ H&C). Since odds are an increasing function of probabilities, the usual Bayesian criterion for the confirmation of H by E given C (i.e., P(H|E&C) > P(H|C)) implies that E confirms H given C just in case Odds(H|E&C) > Odds(H|C), which, by (1), holds just in case LR > 1, assuming nonzero prior odds. Equation (1) makes evident the central role that LR plays in Bayesian confirmation theory generally, since not

8. One can always specify lower (or upper) bounds, but greatest lower (or least upper) bounds arguably face the same problem as point values themselves.

https://doi.org/10.1086/693466 Published online by Cambridge University Press

only does it supply a simple criterion for whether E confirms H, but it also shows exactly how much the prior odds get raised. Odds and probabilities are equally legitimate measures of uncertainty, but in the context of confirmation theory the odds scale has the advantage in making the relation between prior and posterior uncertainties linear with LR furnishing the gradient. Thus, (1) tells us immediately that for a sufficiently large LR all except a small interval $[0,\varepsilon)$ of prior probabilities will generate posterior probabilities/odds greater than any specified value. This fact will be useful later.

With GR substituted for H and M for E, (1) is a simple consequence of the probability axioms and hence true independently of whether P is interpreted according to the subjective or the objective account. We know, however, that in the former, M is part of C, and so of course Odds(GR|M&C) reduces to Odds(GR|C), with the consequence that no Bayesian confirmation of GR by M is possible. The situation is entirely different in the objectivist theory. Here the Jaynesian robot's task is first to compute the probability of GR given information about other relevant parts of theoretical and experimental physics at the time, including information relevant to its prediction of M, and then see what difference to the probability is caused by adding M itself to that information.

Let us for the moment put the first computation on one side and examine the second. Equation (1) tells us that this amounts to evaluating the LR, $P(M|GR\&C)/P(M|\sim GR\&C)$. In the subjective theory, of course, these probabilities are also 'contaminated' with the presumption that M is true, so both numerator and denominator will be equal to 1. But how do we proceed within the objectivist account? The answer is simple in the case of the numerator P(M|GR&C): the observational data (i.e., M) specified Mercury's perihelion advance at $45'' \pm 5''$ per century, and Einstein predicted it to be $43''$, which means that we can reasonably regard M as entailed by GR&C, so that $P(M|GR\&C) = 1$ by the laws of probability alone. As to $P(M|\sim GR\&C)$, we can simplify: since the only seriously entertained alternative at the time to GR was CT, we can to a good enough approximation equate $P(M|\sim GR\&C)$ to P(M|CT&C), giving another simple likelihood so that $P(M|GR\&C)/P(M|CT\&C)$ becomes a true LR.

So far so good. To evaluate P(M|CT&C) itself, remember that in this approach the term on the left-hand side of the vertical stroke depends only on that on the right, with P functioning simply as an inference engine. Thus, here there is no contamination of the probability by M itself, whence P(M|CT&C) can be equated simply to the probability that CT&C themselves assign to M.[9] That the probabilities appearing in statistical models determine rational expectations seems to have been assumed from the early eighteenth century on-

9. This coheres with the fact that if H&C entails E then $P(E|H\&C) = 1$: H&C asserts that E will occur with a probability of 100%.

https://doi.org/10.1086/693466 Published online by Cambridge University Press

ward, and all applied science and all societal infrastructure employing those models assumes such interpretability (in the extreme case, casino operators).[10] Indeed, the rule has seemed sufficiently obvious to generations of Bayesians to be invoked without comment, starting with the eponymous founder himself and continuing through to the present.[11] In Jeffreys's seminal work (1961) the likelihoods are called "direct probabilities,"[12] a terminology Hawthorne repeats in calling the likelihoods "direct inference" likelihoods (2005, 286).

To continue, we know that M was strongly at variance with the prediction of CT, which in conjunction with the known probable error would assign the advance an extremely small probability,[13] whence in accordance with the rule just enunciated we set P(M|CT&C) equal to that same very small probability. Thus, the very small value of P(M|CT&C) implies that we have a very large LR in favor of GR. That by itself, however, does not tell us anything about how large a confirmation, in terms of incremental probability, M accords GR. It is obviously true that for any given positive ε, no matter how small, a value of the LR exists such that for all prior odds greater than ε the posterior odds will exceed any value specified in advance.[14] But the other side of the coin is that however large the LR is, for any given positive ε, all prior odds less than ε/LR will give posterior odds less than ε. Perhaps, though, we can circumvent the need for priors at all by simply taking the LR itself as a standalone measure of confirmation or some suitably increasing function of it like the log to any base greater than 1 (logging means that the measure adds over conditionally independent pieces of conjoined evidence). Taking the name from Peirce, Good (1983, 159) called $\log_{10}$LR "weight of

10. As the previous footnote indicates, this does not exclude these probabilities from being parameters in physical or biological systems like statistical and quantum physics, genetics, etc.

11. In his celebrated calculation of the posterior distribution of a binomial parameter, Bayes simply set the likelihood $P(S_n = r|B(n, p) \,\&\, Q)$ equal to ${}^{n}C_{r}p^{r}(1 - p)^{n-r}$, where B(n,p) states that the sample is drawn from a Bernoulli process of length n with probability parameter p, S_n is the number of successes in the sample, and Q describes the experimental setup, a uniformly smooth, level billiard table with balls set to roll a random distance parallel to one side (the notation is not of course Bayes's own).

12. These are "a pure matter of inference from the hypothesis to the probabilities of different events" (Jeffreys 1961, 57).

13. "Calculation gives for the planet Mercury a rotation of the orbit of 43″ per century, corresponding exactly to astronomical observation. . . . The astronomers have discovered in the motion of the perihelion of this planet, after allowing for all disturbances by other planets, an inexplicable remainder of this magnitude" (Einstein 1916/1952, 200).

14. I hope I will be forgiven for occasionally switching back and forth in midtext between odds and probabilities; they differ only in choice of scale, and sometimes it is more convenient to refer to one rather than the other.

https://doi.org/10.1086/693466 Published online by Cambridge University Press

evidence" (but as we will see below, he added a crucial caveat), suggesting that it does furnish such a measure of confirmation. That way we still get the desired conclusion—that M strongly confirms GR—without having to worry about prior odds at all.

Although the idea has had influential advocates (it is one of several candidate measures of evidential support studied by Eells and Fitelson [2001]; they awarded it top equal marks with the measure $P(H|E) - P(H)$), it is vulnerable to easily manufactured counterexamples. Suppose that a fair-looking coin is tossed 100 times and the observed frequency f is within the 95% probability interval conditional on H, the independent, identically distributed hypothesis. Let E state the value of f and let H' be the conjunction of E and 'most of the moon is made of green cheese': $P(E|H)$ is extremely small, while $P(E|H') = 1$, giving an astronomical LR in favor of H'. But nobody in his or her right mind (presumably) would say that H' is enormously confirmed by E, particularly if, as seems reasonable, H' is assigned prior odds of 0; on the contrary, the verdict should be that E gives H' no support whatever but supports H quite well. On the zero prior assignment, Bayes's Theorem correctly gives H' zero posterior odds however large the LR might be. One cannot object that the likelihood for a hypothesis with zero prior probability/odds is undefined, for it is well known that within extant axiomatizations of conditional probability zero-probability propositions can be consistently conditionalized so long as they are not formal contradictions. Nor should we agree with the demand sometimes made that only formal contradictions be assigned probability 0 (this is the doctrine called Regularity), since it is (*a*) arbitrary—there seems no good reason to deny probability 0 to certainly false contingent sentences, like the one in the example—and (*b*) inconsistent with the ubiquitous employment of continuous distributions in mathematical probability and statistics. In continuum-sized outcome spaces all but countably many possibilities must be assigned probability 0. In any case the example is scarcely weakened if H' is assigned a minuscule positive probability.[15] We need priors after all.

So how should we evaluate the prior odds/probability on GR? GR was and still is almost universally reckoned to be a theory of great simplicity, organically unified around a powerful heuristic incorporating Mach's Principle, the Equivalence Principle, general covariance, the need to deliver Newton's

15. Or even infinitesimal positive probability: such assignments are now known to be perfectly consistent relative to the first-order theory of the real numbers. They are an elementary extension of the reals. Some people have appealed (I think unsuccessfully) to infinitesimals in an attempt to save Regularity from objection *b* (Howson 2016). Compare Good, on the evidential support of the Schrödinger equation: "The large weight of evidence [logLR] makes it seem, to people who do not stop to think, that the initial probability of the equation . . . is irrelevant; but really there has to be an implicit judgment that the initial probability is not too low" (he suggested a threshold of 10^{-50}; 1983, 37).

https://doi.org/10.1086/693466 Published online by Cambridge University Press

gravitation law in weak gravitational fields, and using no special assumptions to predict M and the other early successes. Moreover, since the Jaynesian robot's standards are our own idealized standards built into its probability function, we might in principle simply let the prior odds reflect the opinion of those whose own reasoning is usually supposed to approach most closely the ideal, that is to say the expert opinion represented by the upper end of the scientific hierarchy (e.g., Salmon 1990, 182). While not all were at first as enthusiastic about its intrinsic virtues as Einstein himself and others like Eddington (and later Dirac),[16] the speed with which almost the entire physics community swung behind GR after the early tests would have been very unlikely lacking any confidence in the fundamental theory (to take a well-known example, Dirac all his life thought that the existing theory of quantum electrodynamics was fundamentally wrong despite its "excessively good agreement" with experiment, i.e., predicting the magnetic dipole moment of the electron to an accuracy of 11 decimal places; Dirac 1984, 66).[17]

These observations seem (more than) enough to justify endowing GR with a prior that need not be further specified—which anyway seems hardly possible—to allow us to reasonably conclude that GR was very strongly confirmed by M even though M was old evidence (a fact that has clearly been irrelevant to that calculation).[18] Good, considering the impact not of M but of the gravitational deflection of light, claimed that "General Relativity has very heavy odds on as compared with Newtonian physics," noting that in the calculation he had employed "extravagantly large" initial odds on Newtonian physics to make the point just how little, because of the magnitude of the LR, such inaccuracy influences the posterior odds (1983, 161).

In Hartmann-Fitelson language (see n. 1) we can conclude that GR is an eminently satisfactory explanation of M compared with that provided by CT, with satisfactoriness rendered in a straightforwardly Bayesian way as the possession of a very large LR in its favor combined with a nonnegligible prior whose exact value is irrelevant. Thus, we have a classical Bayesian explanation not only of why Glymour, Garber, and others were correct in highlighting the role played by the prediction of M in very strongly con-

16. "The Einstein theory of gravitation has a character of excellence of its own. Anyone who appreciates the fundamental harmony connecting the way Nature runs and general mathematical principles must feel that a theory with the beauty and elegance of Einstein's theory has to be substantially correct" (Dirac 1978).

17. No-Miracles Argument enthusiasts might ponder this. It was pointed out in Howson (2000, 56–58) that the condition for that argument to be probabilistically valid is that the prior odds be not too small.

18. As Rosenkrantz (1983, 85) noted, we have an explicit expression for the probability of M itself given C: by the theorem of total probability, $P(M|C) = P(M|GR\&C)P(GR|C) + P(M|CT\&C)P(CT|C) = p + \varepsilon(1 - p)$, where $p = P(GR|C)$ and ε is very small, whence we infer that even granted the indeterminate nature of p, P(M|C) is certainly less than 1.

https://doi.org/10.1086/693466 Published online by Cambridge University Press

firming GR but also of why they were wrong in denying that it was M itself that was the confirming agent. I suggested earlier that any situation in which expert opinion decides, as seems par excellence to be the case with M and GR, that a piece of evidence powerfully confirms some given scientific theory can plausibly be seen as a test case for any theory of confirmation, to see whether it succeeds in modeling that opinion without undue difficulty. Apart from ruling out Garber-style accounts, this desideratum also rules out a Bayesian confirmation theory based on subjective probabilities because of the old evidence problem. By contrast, the Objective Bayesian account seems to succeed very well.

There is a large overlap between the Objective Bayesian solution to the old evidence problem and one offered recently by Hawthorne (2005), in which he develops a theory of what he calls agent-based "support functions," whose posterior evaluations ("posterior plausibilities") are based on the sorts of "public" quasi-logical likelihoods central to the Objective theory, combined with priors required only to discount any evidence for them that introduces a prior bias for or against the hypothesis whose posterior plausibility is to be evaluated (305). As I noted earlier, Hawthorne even uses the phrase "direct inference likelihoods," echoing the great Objective Bayesian Harold Jeffreys who called them "direct probabilities," yet Hawthorne never mentions Jeffreys or the Objective theory or Rosenkrantz's demonstration 30 years earlier that it solves the old evidence problem using standard Bayesian techniques. Worse, Hawthorne's account does not actually solve the old evidence problem, which was to explain in Bayesian terms why informed scientific opinion should believe that M very powerfully confirmed GR. Hawthorne's theory does not do this precisely at the point it departs from the Objective theory, in the freedom Hawthorne allows in the choice of priors. To take an admittedly extreme case, a follower of Popper who believes that the priors on all general scientific hypotheses should be set at 0 could still be a Hawthornian, evaluating the posterior support of GR at 0 on M. Of course, the priors can be set to give a more appropriate result, but parameter fixing to save the phenomena hardly counts as an explanation. If Garber fails the explanatory test here, as I have argued, so does Hawthorne. And it is not clear why, as Hawthorne claims, Bayesians also need a subjective belief function when, as he justly points out, constrained only by coherence both the priors and the likelihoods are otherwise free parameters. It is true that he requires the belief functions to be 'aligned' with the support functions, but that makes the latter fundamental and the former effectively (or perhaps ineffectively) otiose.

5. Loose Ends. Some loose ends remain to be tied: three to be precise. The first concerns logic, more specifically logic in the context of Bayesian probability. According to its principal authors Frank Ramsey and Bruno de Finetti, the subjective Bayesian theory should also be subsumed under the

https://doi.org/10.1086/693466 Published online by Cambridge University Press

heading of logic, by which they meant a logic of probabilistic consistency, according to which if this is your degree of belief at time t, then that must also be on pain of inconsistency (Ramsey 1926/1931, 184; de Finetti 1974, 215).[19] In this logic, probability assignments function as premises, from which deductive logic plus the probability calculus elicits consequences. Now, however, we seem to have a problem: if the subjective theory also has a logical character, then it might seem that anything the Objective Bayesian theory can do in the way of presenting a Bayes's Theorem inference *sub specie aeternitatis*, as it were, the subjective theory can do likewise—even if the appearance might differ. In principle, then, it seems that it might solve the old evidence problem after all: simply adopt as a premise that $P(M) < 1$ and infer within this logic that $P(GR|M) > P(GR)$.[20] Sadly no: the inference is valid but not sound, since for a subjective Bayesian the premise $P(M) < 1$ is incoherent: a basic principle of the subjective theory is that your subjective probability function must reflect your true beliefs at the time; otherwise, when penalized by a proper scoring rule for inaccuracy in your forecasting of random events (de Finetti 1974, 90), you will do better, with regard to expectations, to change it to one that does reflect that belief.[21]

The second loose end is what Garber called 'logical omniscience'. One of the probability axioms is that every logical truth receives probability 1. To be able to decide logical truth for arbitrary sentences in anything stronger than propositional logic is something that Church's Theorem tells us not even a Universal Turing Machine can do. Hence, talk of these probabilities being 'degrees of rational belief', or 'rational expectations', might appear to make any theory of epistemic probability based on the usual axioms vulnerable to the charge of being hopelessly unrealistic. Fortunately, that fear is groundless: 'logical omniscience' is no more problematic for this theory than it is for deductive logic itself (where no one suggests that it should be weakened to take account of Church's result). We may not be able to decide whether an arbitrary sentence is a logical truth, but we can be taught to recognize some logical truths, usually by proving them or having them proved for us within some accepted deductive system. And we can and do tie our notion of rationality to the capacity to be convinced by arguments shown to meet the canons

19. In English translations of Finetti's later work, "consistency" has been supplanted by "coherence," but de Finetti certainly regarded his theory as a species of logic (actually, multivalued logic). Howson (2000, 127–34) argues that there is an interesting kinship between probabilistic and deductive consistency (both exemplify the mathematical idea of consistency as solvability subject to constraints).

20. We see later that this is the strategy of the so-called counterfactual solution of the old evidence problem.

21. De Finetti's scoring rule, a version of the well-known Brier rule, is equal to minus the mean-squared error. A proper scoring rule, of which that is one, minimizes your expected penalty just when your estimates reflect your true beliefs.

https://doi.org/10.1086/693466 Published online by Cambridge University Press

of logical deduction, as the drive for logical rigor and the development of provably sound deductive systems for mathematics over the last century attests. It is of course true, as Garber so famously pointed out, that the learning of a logico-mathematical truth cannot confirm any hypothesis in any theory based on the usual probability axioms, but if the foregoing is correct that is irrelevant to the solution of the old evidence problem.

Finally we need to see how, or for that matter whether, the solution offered above escapes the objections brought against Objective Bayesianism in an influential work on Bayesian confirmation theory by Earman (1992). Earman voices two general objections, neither of which seems very compelling. The first concerns familiar difficulties in the way of 'objectifying' prior probabilities. It is indeed true, as I pointed out earlier, that the usual attempts to do so have hardly met with unalloyed success, but I argued that where the LR is very large the problem can be largely avoided by exploiting the robustness of the posterior odds under very large variations in the prior and assuming, as seemed very reasonable in the case of GR, a nonnegligible but otherwise undetermined prior probability. To the objection that this does not extend to cases in which the LR is not so large, my answer is that it actually does, although generally at the cost of some degree of indeterminacy. But just as admitting priors is, I believe, an unavoidable feature of sound inductive inference, equally unavoidable in my opinion is indeterminacy, sometimes more, sometimes less, and sometimes unfortunately quite a lot. But probably that is just how it should be: things are not always precise, and where they are not it is usually a mistake to try to make them so. Garbage in, as they say, garbage out.

Earman's other objection is that if the probability in question is "interpreted as degree of belief, rational or otherwise, then it must be time-indexed" (1992, 120). No argument is provided for this judgment, which is contradicted by the fact that under the objectivist interpretation the probability relation (described by Cox as determining degree of rational expectation) is not of this type and also by Earman's own exercise in Bayesian confirmation theory applied to Hume's famous argument against miracles, where he says this: "My proposal starts from the fact that Hume describes a situation in which it is known that the witness has testified to the occurrence of a miraculous event. *Thus, we should be working with probabilities conditioned on t(M)* [i.e., the testimony that M is true, where M is the claim that the miracle occurred], as well as on the evidence of experience [E] and the other background knowledge K. In such a setting, the probability of the event the testimony endeavours to establish is $\Pr(M|t(M) \,\&\, E\&K)$" (Earman 2000, 41; my emphasis).

6. 'Counterfactual' and Related Approaches. That Earman in the quotation above is ostensibly appealing to the subjective Bayesian theory, yet

https://doi.org/10.1086/693466 Published online by Cambridge University Press

he assigns to old evidence a probability less than 1 given other background information, seems to be an example of what has become known as the 'counterfactual' solution of the old evidence problem (even though Earman [1992] is dismissive of it). According to that account, for the purpose of assessing the confirmation of a new theory H by old evidence E, E is simply dropped, to the extent that doing so is possible, from the body of the agent's total knowledge even though the probability function is supposedly a subjective one. Indeed, Howson and Urbach's well-known subjective Bayesian tract (2006) defends the counterfactual view, proposing the same less-than-unity total probability evaluation of P(M|C) as P(M|GR)P(GR) + P(M|CT)P(CT) as in the Objective theory, with the likelihoods 1 and a small number, respectively.

In defense of this position, one might observe that the word 'prior' in 'prior probabilities' is usually understood to mean 'prior to the evidence', and when the evidence is already there it seems a natural move to see whether it can somehow be kept separate from the rest of the information on which one bases one's beliefs. Unfortunately for naturalness, there are two reasons why in the context of the subjective theory this is not a feasible strategy. First, as has often been pointed out, the attempt to shield the measure of confirmation from evidence already known is not always possible: in certain cases that evidence may be too entangled with other parts of the background information to be cleanly extracted from it (Howson and Urbach [2006] agree). Second, it is actually inconsistent, or incoherent in standard subjective Bayesian language, since as I pointed out earlier, under the rules of that theory you are compelled to assign probability 1 to all your current beliefs. In effect the counterfactual strategy is an attempt to graft Objective probability evaluations onto a subjectivist stock, and apart from being rather obviously ad hoc it does not work.

A recent development of the counterfactual approach to the old evidence problem—although the authors do not describe it as such—is developed in Romeijn and Wenmackers (2016), although the sort of old evidence they are principally concerned with is that which suggests that a new hypothesis, H, provides a better account of it than those initially considered, and their objective is to explain how what they call 'standard Bayesianism' can be adapted to deal with such a situation. In the course of their discussion, they argue that in computing the degree of confirmation of H by such evidence E, E may be assigned a probability less than 1 even though it was known and indeed instrumental in generating H. The authors claim that this strategy also solves the historical old evidence problems (presumably including the Mercury one even though GR was not constructed to accommodate the anomalous orbital motion), but in that case their approach is practically indistinguishable from the counterfactual one and, hence, also vulnerable to the same charge of adhocness: they agree that what they are doing might be described as "reverse-engineering" the priors (1242), yet they also claim that their time-indexed

https://doi.org/10.1086/693466 Published online by Cambridge University Press

probability function represents the "rational degree of belief of an agent" (1230). But the rational degree of belief in E of an agent who already knows E is 1 (the Objective account, also nominally one of rational degree of belief, can legitimately assign E a probability less than 1 because it is relative to specified information). I think the verdict must be that within the formal and interpretive confines of Romeijn and Wenmackers' theory the historical old evidence problem remains unsolved.

7. Conclusion. The subjective Bayesian theory as developed, for example, by Savage in his classic text (1954; Savage calls it "personal probability") may justly claim for itself a rigorous development from first principles,[22] but it cannot solve the deceptively simple but actually intractable old evidence problem, whence as a foundation for a logic of confirmation at any rate, it must be accounted a failure. I argued in the first part of this article that the nonclassical developments of that theory initiated by Garber in an attempt to solve the problem fail too: in addition to experiencing severe and I believe insuperable internal problems they cannot explain why, according to practically all expert opinion including that of Einstein himself, Mercury's anomalous orbital motion was an extremely strong confirmation of GR, and we have seen Garber himself implicitly deny it. By contrast, the Objective Bayesian approach has no difficulty in accommodating it in a simple Bayes's Theorem computation.

This is not to say that Garber's theory and the subsequent developments it has generated are of no interest: that is far from true, and Garber highlighted a problem that all varieties of Bayesianism must at some point confront, which is how it can be extended to pure-mathematical theories (and this includes questions of logical derivability since these can be encoded within set theory and indeed, if they concern first-order logic, within arithmetic). These theories are after all factual propositions of a sort (I am assuming without argument that mathematical logicism is to all intents and purposes dead), and mathematicians certainly talk about the probability that some conjecture or other, like the Riemann hypothesis, is true given evidence consisting of a very large number of 'confirming' instances. To put this in good Bayesian order is indeed a pressing need (for a comprehensive discussion of the problems facing such an account, see Corfield [2001]). That said, I fear that in Garber's claim, quoted earlier, that "if old evidence can be used to *raise* the probability of a new hypothesis, then it must be by way of the *discovery* of previously unknown logical relations. In the cases that give rise to the problem of old evidence, we are thus dealing with circumstances in which hypotheses are confirmed not by the empirical evidence itself, but by the discovery . . . that

22. Although as is well known some of these, like the independence principle, have been seriously questioned.

https://doi.org/10.1086/693466 Published online by Cambridge University Press

$h \vdash e$," we merely have a non sequitur generating a minor industry in proving something that is not true.

REFERENCES

Corfield, David 2001. "Bayesianism in Mathematics." In *Foundations of Bayesianism*, ed. D. Corfield and J. Williamson, 175–203. Dordrecht: Kluwer.

Cox, Richard T. 1946. "Probability, Frequency and Reasonable Expectation." *American Journal of Physics* 14:1–13.

———. 1976. *Of Inference and Inquiry: An Essay in Inductive Logic*. Baltimore: Johns Hopkins University Press.

de Finetti, Bruno. 1974. *Theory of Probability*. Vol. 1. Wiley.

Dirac, Paul A. M. 1978. "The Excellence of Einstein's Theory of Gravitation." Paper presented at the Symposium on the Impact of Modern Scientific Ideas on Society, Munich, September 18–20. http://unesdoc.unesco.org/images/0003/000311/031102eb.pdf.

———. 1984. "The Requirements of Fundamental Physical Theory." *European Journal of Physics* 5:65–67.

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

———. 2000. *Hume's Abject Failure: The Argument against Miracles*. Oxford: Oxford University Press.

Eells, Ellery. 1990. "Bayesian Problems of Old Evidence." In *Minnesota Studies in the Philosophy of Science*, ed. C. Wade Savage, 224–45. Minneapolis: University of Minnesota Press.

Eells, Ellery, and Branden Fitelson. 2001. "Symmetries and Asymmetries in Evidential Support." *Philosophical Studies* 107:129–42.

Einstein, Albert. 1915. "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie" [Explanation of the perihelion motion of Mercury from general relativity theory]. In *Sitzungsberichte der Königlich Preussische Akademie der Wissenschaften*, 831–39. Berlin: Königlich Preussische Akademie der Wissenschaften.

———. 1916/1952. "The Foundation of the General Theory of Relativity." In *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity*, ed. H. A. Lorentz, Albert Einstein, Hermann Minkowski, and Hermann Weyl, 109–64. New York: Dover.

Garber, Daniel. 1983. "Old Evidence and Logical Omniscience in Bayesian Confirmation Theory." In *Minnesota Studies in the Philosophy of Science*, ed. J. Earman, 99–131. Minneapolis: University of Minnesota Press.

Glymour, Clark. 1980. *Theory and Evidence*. Princeton, NJ: Princeton University Press.

Good, Irving J. 1983. *Good Thinking*. Minneapolis: University of Minnesota Press.

Goodman, Nelson. 1946. *Fact, Fiction and Forecast*. Indianapolis: Bobbs-Merrill.

Halpern, Joseph Y. 1996. "A Counterexample to Theorems of Cox and Fine." In *Proceedings of the AAAI Conference*, 1313–19. Menlo Park, CA: AAAI.

Hartmann, Stephan, and Branden Fitelson. 2015. "A New Garber-Style Solution to the Problem of Old Evidence." *Philosophy of Science* 82 (4): 712–17.

Hawthorne, James. 2005. "Degree-of-Belief and Degree-of-Support: Why Bayesians Need Both Notions." *Mind* 114:277–30.

Howson, Colin. 2000. *Hume's Problem*. Oxford: Clarendon.

———. 2016. "Repelling a Prussian Charge with the Solution to a Paradox of Dubins." *Synthese*. doi:10.1007/s11229-016-1205-y.

Howson, Colin, and Peter Urbach. 2006. *Scientific Reasoning: The Bayesian Approach*. 3rd ed. Chicago: Open Court.

Jaynes, Edwin T. 1973. "The Well-Posed Problem." *Foundations of Physics* 3:477–93.

———. 2003. *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.

Jeffreys, Harold. 1961. *Theory of Probability*. 3rd ed. Oxford: Oxford University Press.

Ramsey, Frank P. 1926/1931. "Truth and Probability." In *The Foundations of Mathematics*, ed. Robert B. Braithwaite, 156–98. London: Kegan Paul, Trench, Trubner.

https://doi.org/10.1086/693466 Published online by Cambridge University Press

Romeijn, Jan-Willem, and Sylvia Wenmackers. 2016. "A New Theory about Old Evidence: A Framework for Open-Minded Bayesianism." *Synthese* 193:1225–50.
Rosenkrantz, Roger D. 1983. "Why Glymour Is a Bayesian." In *Minnesota Studies in the Philosophy of Science*, ed. J. Earman, 69–97. Minneapolis: University of Minnesota Press.
Salmon, Wesley. 1990. "Tom Kuhn Meets Tom Bayes." In *Scientific Theories*, ed. C. Wade Savage, 175–204. Minneapolis: University of Minnesota Press.
Savage, L. J. 1954. *The Foundations of Statistics*. London: Wiley.
Snow, Paul. 2001. "The Disappearance of Equation Fifteen, a Richard Cox Mystery." In *Proceedings of the Fourteenth International Florida Artificial Intelligence Research Society Conference*, ed. Ingrid Russell and John Kolen. Menlo Park, CA: AAAI.
Terenin, Alexander, and David Draper. 2015. "Rigorizing and Extending the Cox-Jaynes Derivation of Probability: Implications for Statistical Practice." ResearchGate.net.

https://doi.org/10.1086/693466 Published online by Cambridge University Press