THE JOURNAL OF PHILOSOPHY

VOLUME XCVI, NO. 9, SEPTEMBER 1999

MEASURING CONFIRMATION*

When belief comes to be seen as involving degrees of confidence rather than simple yes-or-no judgments, the theory of rational belief must be adjusted accordingly. Probabilistic coherence stands out as an attractive analogue to the traditional desideratum of deductive consistency. And probability theory, as developed in Bayesian theories of rational belief, seems to go beyond this. It provides a basis for conditionalization principles regulating change of belief—a topic about which traditional logic had little to say; and it offers a quantitative analysis of our notion of confirmation, or evidential support. Confirmation, like belief itself, is something that obviously comes in degrees, but formal confirmation models based on deductive logic are merely qualitative. It seems that the relation these accounts seek to model must derive from a deeper, quantitative one; and Bayesians seem to be in a good position to provide one. According to the standard "positive relevance" account offered by Bayesians, a bit of evidence E confirms hypothesis H to the extent that it makes H more probable, in the sense that $\mathrm{pr}(H/E) > \mathrm{pr}(H)$.

Now, this analysis cannot be thought to match our common-sense concept of evidence exactly. For one thing, the probabilistic relation of positive relevance is symmetrical: E will confirm H on this analysis just in case H confirms E. But our ordinary concept of evidence involves a certain kind of epistemic asymmetry. Hypotheses are commonly taken to be supported by evidence, but not vice versa.

* I would like to thank Jim Joyce, Mark Kaplan, Hilary Kornblith, Don Loeb, Patrick Maher, Derk Pereboom, and Lyle Zynda for helpful discussions, correspondence, and/or comments on earlier drafts. Thanks also to the University of Vermont for research support.

0022-362X/99/9609/437–61

We do not commonly say, in cases where E is evidence for H, that H is also (at least some) evidence for E.

Moreover, our common-sense concept may well be indeterminate, in a way that will necessarily fail to match up with any precise formal model. Some have argued that the standard model, which is based on the *difference* between $pr(H/E)$ and $pr(H)$, should be replaced by a model based on the *ratio* of these quantities, or on the logarithm of this ratio.[1] In support of such suggestions, it has been pointed out that, for instance, a bit of evidence that raised the probability of a hypothesis from $1/10^9$ to $1/100$ would intuitively seem much more momentous than one that raised the probability of a hypothesis from .26 to .27.[2] On the ratio-based model, but not on the difference-based model, the former bit of evidence would be seen as more confirmatory. But there are intuitions on the other side here, too. If we compared a bit of evidence that raised the probability of a hypothesis from .0001 to .001, it would likely strike us as much less powerful than a bit of evidence that raised a hypothesis from .1 to .9. This supports the difference model over the ratio model.[3] What is a confirmation theorist to do? One might seek to settle the dispute in favor of one of the two measures by supporting some of these intuitions and/or undermining others. Or one might reject both measures, and try to accommodate all the intuitions in some third measure. But, to my mind, it is not at all clear that this sort of intuitive conflict should be seen as demanding formal solution. One may instead see it as a symptom of a relatively straightforward indeterminateness of the ordinary concept of evidential support.

Compare, for example, the notion of financial support. Suppose we are asked how we should measure the degree to which a candidate C is supported by an interest group I. One way of measuring this relation would be in terms of the number of dollars C received from I. But this measure would miss some important aspects of the situation. We might ask what proportion of C's funds are supplied by I. Or, looking at the matter from a third angle, we might ask how crucial I's support is for C (it might not be crucial even if I is a high-dollars, high-percentage contributor, if C has sufficient other funds). Or we could use still another measure and ask what C's position would be if I were the only contributor. Thinking about these different measures of support suggests to me that there is no single clear-cut question being asked when we ask 'How much support does C get from I?' It would not be surprising if the same were true of the question 'How much does evidence E support hypothesis H?'

Nevertheless, despite the fact that it fails to capture epistemic asymmetry, and despite the apparent indeterminateness of our pretheoretic notion, the positive-relevance model seems to be reflecting something important. Surely, 'making more probable' corresponds to something central in our notion of evidential support. The model has been vigorously employed in Bayesian explanations of various tenets of scientific (and common-sense) epistemology. It would, I think, be surprising if some notion along the lines of positive relevance were not at least an important component of our natural notion of evidence.[4]

I. QUANTITATIVE CONFIRMATION AND OLD EVIDENCE

Unfortunately, this model (along with the ratio- and log-of-ratio-based variants) seems to have a fatal flaw. The problem first emerges when considering cases in which $pr(E)$ is taken to have the extreme value of 1. In those circumstances, E is treated probabilistically just as if it were a tautology. It cannot confirm anything, since $pr(H/E)$ will be equal to $pr(H)$. Once one is absolutely certain about one's evidence, it no longer is evidence.[5] Unlike the asymmetry and indeterminateness problems, the problem of old evidence threatens to vitiate the probabilistic approach to confirmation.

Against this version of the problem, it may be protested—reasonably enough, by common Bayesian standards—that we should never believe a nontautology with probability 1 anyway. But this response does not remove the problem's sting. For it remains true that as $pr(E)$ approaches 1, the degree to which E can confirm anything becomes vanishingly small. The more confident we become in our evidence, the less it can *be* evidence.

This quantitative version of the old-evidence problem is not restricted to the standard relevance measure. It also infects the ratio-

[1] For the former suggestion, see George N. Schlesinger, "Measuring Degrees of Confirmation," *Analysis*, LV, 3 (1995): 208-12; for the latter, see Peter Milne, "A Bayesian Defence of Popperian Science?" *Analysis*, LV, 3 (1995): 213-15.

[2] See Schlesinger, pp. 210-11.

[3] See Elliott Sober, "No Model, No Inference: A Bayesian Primer on the Grue Problem," in Douglas Stalker, ed., *Grue! The New Riddle of Induction* (Chicago: Open Court, 1994), p. 228, where the example is credited to Ellery Eells.

[4] What the natural intuitive notion is may not be entirely clear. The core of the intuitive notion I wish to explore here takes confirmation as a relation of support between propositions which explains why confidence in the supporting proposition can help make rational confidence in the supported proposition, where the confidence involved comes in degrees. One might alternatively see confirmation as most fundamentally aimed at explaining rational *acceptance* of propositions. If one took this latter approach, the issues discussed in this paper might look quite different. See, for example, Isaac Levi, *The Enterprise of Knowledge* (Cambridge: MIT, 1980).

[5] Credit for pointing out this problem goes to Clark Glymour; see his *Theory and Evidence* (Princeton: University Press, 1980), pp. 85-93.

and log-of-ratio-based variants mentioned above. As John Earman points out, the problem also infects Haim Gaifman's confirmation measure (1-pr(H))/(1-pr(H/E)).[6] What these measures all have in common is that they ultimately root confirmation in the contrast between pr(H) and pr(H/E). As pr(E) nears 1, this contrast disappears.

A somewhat different probabilistic measure that has been championed by I. J. Good[7] is different in this respect; it is based on contrasting the *likelihoods* pr(E/H) and pr($E/{\sim}H$). He calls the ratio of these quantities "the (Bayes) factor in favor of H provided by E," and the logarithm of this ratio "the weight of evidence in favor of H provided by E." The two likelihoods compared in Good's measure do not converge as pr(E) approaches 1 in quite the same way as pr(H) and pr(H/E) do. Nevertheless, this measure, too, turns out to be infected by quantitative old-evidence difficulties. In cases where E confirms H, the likelihood ratio ranges between 1 and ∞. But when pr(E) is high, and pr(H) is moderate—perhaps the paradigmatic situation in which one discusses evidence—the likelihoods are forced to be so close that their ratio falls almost all the way toward the minimum end of this spectrum.[8]

One response to this pervasive problem is simply to give up on finding a Bayesian analysis of our quantitative notion of confirmation. Suppose we discount the cases where pr(E)= 1, on the grounds that a model of ideal rationality need not account for cases where agents give probability 1 to nontautologies. Once we do this, we are left with an attractive analysis of *qualitative* confirmation, for pr(H/E) can still be a tiny bit higher than pr(H) when pr(E) is very close to 1.[9] But many Bayesians have been reluctant to take this line. A literature has grown up around trying to solve the problem of old evidence.

One of the accomplishments of this literature has been to identify different problems of old evidence. Some detailed taxonomies have been offered, but here I chiefly want to distinguish two species of problem. One may think of the problem, as Clark Glymour (*op. cit.*) seems to have, as a problem for the Bayesian theory of change of belief. The problem here is that scientists seem to use evidence of which they have long been confident to *raise* the probability of a hypothesis. This can happen when the hypothesis is thought up at a point when the evidence is already known, or when someone suddenly realizes that the hypothesis entails or explains the evidence. Most of the old-evidence literature is primarily focused on this sort of difficulty for the Bayesian account of belief change; let us call this the *diachronic problem of old evidence.*

The other sort of old-evidence problem does not (at least directly) involve Bayesian accounts of belief change. 'Confirmation', after all, is not just the name of a kind of event. It also seems to refer to a certain relation between propositions. Some propositions seem to help make it rational to believe other propositions. When our current confidence in E helps make rational our current confidence in H, we say that E confirms H. For such a relation to obtain, no event consisting of our becoming more confident in H on the basis of our confidence in E need be occurring.

Indeed, the focus of traditional confirmation theory abstracted completely from particular believings. For the confirmation relation to hold between E and H, it was not required that E or H actually be believed at all. The confirmational relations were determined by logical relations; the connection to belief was just that, when E confirmed H, belief in E *would* help make belief in H rational.[10] Now, contemporary Bayesians—at least those who have given up Rudolf Carnap's hope of finding the objectively rational probabilistic relations between propositions[11]—cannot so easily separate confirmational relations from relations involving an agent's degrees of belief in propositions. Still, there seems to be a clear sense in which, given the beliefs of a particular agent, certain propositions support others

[6] Gaifman, "On Inductive Support and Some Recent Tricks," *Erkenntnis*, XXII (1985): 5-21; and Earman, *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory* (Cambridge: MIT, 1992), p. 121, fn. 8.

[7] See, for example, "Some Logic and History of Hypothesis Testing," in his *Good Thinking: The Foundations of Probability and Its Applications* (Minneapolis: Minnesota UP, 1983). Thanks to Patrick Maher for prompting me to consider Good's measure.

[8] For example, let pr(E) be .99. If pr(H) is .75, the highest value the ratio can have is about 1.04. (To see this, note that in general, pr(E) = pr(H)pr(E/H) + pr($\sim H$)pr($E/{\sim}H$). So in the present case, we get: .99 = .75pr(E/H) + .25pr($E/{\sim}H$). From this, it is clear that as pr(E/H) gets higher, pr($E/{\sim}H$) gets lower, and that the ratio of pr(E/H) to pr($E/{\sim}H$) will be greatest when pr(E/H) is as large as possible—that is, when it is 1. Assuming the ratio is thus maximized, we get: .99 = .75 + .25pr($E/{\sim}H$), so pr($E/{\sim}H$) = .96. Thus, the maximum value the likelihood ratio can take when pr(E) = .99 and pr(H) = .75 is 1/.96, which is about 1.04.) The value is even lower if we suppose more uncertainty about H; if pr(H) is .6, the ratio maximum drops to under 1.03. And even if we then allow pr(H) to reach .9, the ratio has a maximum of about 1.1.

[9] For a nice defense of this position, see Mark Kaplan's *Decision Theory as Philosophy* (New York: Cambridge, 1996), chapter 2.

[10] See, for example, Carl G. Hempel, "Studies in the Logic of Confirmation," in his *Aspects of Scientific Explanation* (New York: Free Press, 1945), p. 24.

[11] One who has not given up on the objective approach is Maher. In "Subjective and Objective Confirmation," *Philosophy of Science*, LXIII (1996): 149-74, he offers a highly contextualized, yet objective, probabilistic account (though of a purely qualitative concept of confirmation).

to various degrees. The hope is that these confirmational relations (for a given agent at a given time) will be determined by the agent's degrees of belief (which are, of course, logically constrained by the probability axioms).

violation of 01

Unfortunately, analyzing this sort of confirmation in the standard Bayesian way yields the result that my high degree of confidence in E makes E incapable of lending significant rational support to any other belief. Let us call this the *synchronic problem of old evidence*.[12]

II. THE INTEREST OF THE SYNCHRONIC PROBLEM

Various writers have tied the diachronic problem of old evidence to the fact that real agents are imperfect logicians. An agent who is confident of E may fail to see that it follows from H, and then, upon discovering the entailment, raise her confidence in H. This has led to one of the two main lines of approach to the old-evidence problem. The standard notion of probabilistic coherence requires agents to have a certain sort of logical omniscience: they must give probability 1 to all logical truths, and respect all logical entailments perfectly (for example, by never having $\mathrm{pr}(P) > \mathrm{pr}(Q)$ when P entails Q). Once this requirement is relaxed, we might devise ways of describing how learning *that H entails E* can—via probabilistic calculation—raise an agent's confidence in H.[13]

This approach treats the old-evidence problem as a symptom of excessive idealization. Clearly, even the most epistemically perfect among us cannot approach logical omniscience. A theory that builds such omniscience centrally into its epistemic model may lose the ability to give a rational reconstruction of much human behavior that is surely epistemically praiseworthy. Thus, the project of making the Bayesian theory of rationality more realistic is clearly an important one.

[12] My distinction between the diachronic and synchronic problems mirrors Daniel Garber's distinction between the *historical* and *ahistorical* problems, and Lyle Zynda's distinction between the problem of the *confirmation event* and that of the *confirmation relation*. It more roughly corresponds to Eells's distinction between the problem of *new old evidence*, and the problems of *old new evidence* and *old old evidence*. See Garber, "Old Evidence and Logical Omniscience in Bayesian Confirmation Theory," in Earman, ed., *Testing Scientific Theories, Minnesota Studies in the Philosophy of Science*, Volume X (Minneapolis: Minnesota UP, 1983), pp. 99-131; Zynda, "Old Evidence and New Theories," *Philosophical Studies*, LXXVII (1995): 67-95; and Eells, "Problems of Old Evidence," *Pacific Philosophical Quarterly*, LXVI (1985): 283-302.

[13] See Garber; Richard Jeffrey, "Bayesianism with a Human Face," in *Testing Scientific Theories*; and Ilkka Niiniluoto, "Novel Facts and Bayesianism," *British Journal for the Philosophy of Science*, XXXIV (1983): 375-79.

Nevertheless, there is a sense in which the diachronic old-evidence problem, as a symptom of overidealization, leaves an important aspect of the Bayesian model unscathed. It does not show Bayesianism defective as a certain kind of epistemic ideal, or model of rational perfection. It is quite plausible that a being who had unlimited intellectual powers, but limited data, *would* have a probabilistically coherent confidence distribution over all possible hypotheses. For such an ideal being, the diachronic problem of old evidence simply would not arise. This observation at least limits the urgency of the diachronic old-evidence problem, if one is willing to allow for a good deal of idealization in thinking about epistemic norms. And in fact, the degree of idealization involved here should not be too shocking for those accustomed to thinking about ideally rational all-or-nothing beliefs in terms of the consistency and closure conditions of deductive logic.[14]

But the bite of the old-evidence problem goes deeper than revealing the extent of idealization inherent in Bayesianism. This comes out in considering the synchronic version of the problem. Consider an agent who is not only currently logically omniscient, but who has always been so, and who has arrived at her present beliefs by faultless application of the idealized mechanisms of Bayesian belief revision. Even for such an ideal intellect, it seems clear that two things can be true simultaneously: (1) some beliefs will provide significant evidential support for others; and (2) some of the evidential beliefs will be held with high levels of confidence. These two features are, as noted above, flatly incompatible on the standard Bayesian analysis of (synchronic) evidential support. Thus, no lapse from rational perfection, even in the past, is needed for the synchronic old-evidence problem to arise. For this reason, the synchronic problem presents a particularly sharp challenge to the philosophical interest of the Bayesian program: it shows standard Bayesianism defective even as a

[14] It might be objected that one who was comfortable with a lot of idealization would have little reason to prefer Bayesianism over traditional logic in the first place. Is not probability theory itself basically a response to excessive idealization of the all-or-nothing model of belief?

Such a comparison would be misleading. The problem with the bivalent model of belief is not just that it fails to describe actual imperfect human attitudes toward propositions. Even a rationally ideal being would not have all-or-nothing beliefs on contingent matters about which she was incompletely informed. An agent with this limited information who nevertheless harbored only bivalent beliefs would be less epistemically perfect, not just less realistic, than one whose confidence in such propositions came in degrees. Since having only all-or-nothing beliefs based on incomplete information is not a rational ideal, the move to Bayesianism is not one of trading off idealization for realism, but one of improving the idealization.

model of idealized rationality. In what follows, then, I shall focus on the synchronic problem.

III. CAN THE SYNCHRONIC PROBLEM BE REDUCED TO THE DIACHRONIC?

One reason that interest has centered around the diachronic problem may be based on the thought that a solution to the diachronic problem would provide a solution to the synchronic one. Suppose that we had in hand a satisfactory account of confirmation events—events in which probabilities of hypotheses are raised by evidence. A given agent's body of belief will be the result, in part, of a history of such events. Given such a history, Ellery Eells[15] writes that "it seems appropriate to say that *E is (actual) evidence for T*, for a given individual, if, at some time in the past, the event of its confirming *T*, for that individual, took place" (*op. cit.*, p. 287). Earman[16] concurs with this approach, which I shall call the *historical* approach to the synchronic problem.

The historical approach clearly gives up on the idea that confirmational relations (for a given agent at a given time) are determined by the agent's present degrees of belief. Still, there is something attractive about the idea. And if it were successful, the synchronic problem of old evidence would be reduced to the diachronic problem. In that case, the entire old-evidence problem might seem to stem from over-idealization.

It turns out, however, that the historical approach overestimates the connection between the two concepts of confirmation. To see this, consider an agent who is wondering whether deer live in a nearby wood. He comes across a pile of deer droppings, and his confidence in the deer hypothesis increases to near 1. Shortly thereafter, he finds a shed antler. Since his confidence in the deer hypothesis is already so high, this new evidence does not have any significant impact on it. Now, subsequent to the agent's finding the second piece of evidence, let us ask whether the evidence about the droppings and the antler provide rational support for our agent's belief about deer. Intuitively, the droppings and the antler provide equally strong rational support for the agent's deer belief. There is, I think, no sense in which the droppings currently provide stronger support or better evidence than does the antler. But only the droppings are historically associated with a significant increase in the

[15] Eells distinguishes 'confirming' from 'being evidence for', but I shall use the terms interchangeably. His analysis is intended to capture the sense in which *E* supports *H* after *E* is known.

[16] *Bayes or Bust?*, chapter 5.

agent's probability for deer. The historical approach thus makes contemporary evidential support depend in an unintuitive way on the order in which evidence was discovered.[17] Intuitively, synchronic support should depend on the agent's present epistemic state, not on such historical accidents. Clearly, the synchronic problem cannot be reduced to the diachronic one in this way.[18]

IV. THE COUNTERFACTUAL APPROACH

The most immediately appealing, and most widely criticized, direct approach to the synchronic problem relies on counterfactual degrees of belief. For an agent who knows *E*, this approach would analyze confirmation with the standard positive-relevance definition; however, it would be applied not to the agent's actual probabilities, but to what those probabilities would have been in circumstances where the agent did not know *E*.[19]

This approach, like the historical approach, would abandon seeing confirmation relations as determined by an agent's probabilities. Still, the intention seems to be to make confirmation relative to a probability function that is in a sense as close to the agent's as possible, consistent with the agent's not knowing *E*.

[17] I make this argument briefly in a review of Earman's *Bayes or Bust?* in *Philosophical Review*, CIII (1994): 345-47. Maher argues for the same conclusion using a somewhat different type of example. In his example, the bearing of the evidence on a hypothesis changes as background knowledge changes. In such situations, a bit of evidence that intuitively confirms a certain hypothesis today might have disconfirmed it when it was first discovered.

[18] One might object that this sort of problem need not trouble a historical analysis of *qualitative* confirmation. (Earman is clearly interested in a Bayesian analysis of the quantitative notion, but Eells may have intended his argument to apply only to a qualitative notion.) The point would be that finding the antler presumably would have raised the probability of deer at least a tiny bit, given that the probability for deer would always remain a bit shy of unity.

But as mentioned above, it is not obvious that the old-evidence problem arises at all, if we seek only a qualitative analysis of confirmation and preclude complete confidence in contingent matters. In this case, it is not clear that there is any reason to substitute the historical account for the standard, nonhistorical analysis.

On the other hand, if we do allow agents to believe contingent matters with probability 1, then it would be hard to see why the deer hypothesis could not attain this level. And if it did, then the historical account of even qualitative confirmation would fail in just the manner described in the text. Evidence that was discovered subsequent to the hypothesis' being given probability 1 could not increase its probability. Such evidence might nevertheless provide just as much rational support for the hypothesis as would the evidence that historically helped raise its probability.

Finally, the sort of example given by Maher would tell against qualitative versions of the historical account even if we did not allow full belief in contingent propositions.

[19] This approach is tentatively endorsed in Garber and embraced more wholeheartedly in Colin Howson, "Bayesianism and Support by Novel Facts," *British Journal for the Philosophy of Science*, XXXV (1984): 245-51.

But will the counterfactual move allow us to capture what we want to capture from the agent's actual probability function? Various writers[20] have noticed problems here. For one, it is not clear that the counterfactual stipulation determines any unique probability function. Moreover, in certain cases, the distribution that the agent would have in the counterfactual situation would be produced by irrational factors that would accompany the nonbelief in $E2$; in such cases, the counterfactual degrees of belief give the wrong result.[21] These sorts of cases highlight the problem of getting too far away from the agent's real probabilities that were supposed to ground the confirmation relation. The factors that supplement the agent's probabilities in determining the truth of the counterfactuals are often irrelevant to the agent's epistemic state.

IV. DOES JEFFREY CONDITIONING UNDERMINE THE SYNCHRONIC NOTION?

Another approach to the synchronic problem is to argue that, from the Bayesian point of view, the whole project of analyzing synchronic confirmation should be abandoned. Lyle Zynda (*op. cit.*) argues that, once we see the Bayesian position as allowing belief updating by Jeffrey conditioning as well as classical conditionalization, we can see that the whole Bayesian project of analyzing synchronic confirmation falls apart—and for reasons other than the old-evidence problem.[22]

The argument begins by considering cases where $\mathrm{pr}(E)$ is moderate, so the old-evidence problem does not arise. For such cases, Zynda suggests that we should see E as confirming H only if the following condition holds: for any probability function PR that comes from the present probability function pr by Jeffrey conditioning initiated by a set of probability changes which includes raising the probability of E, $\mathrm{PR}(H) > \mathrm{pr}(H)$.

Zynda shows that this condition is not acceptable. An instance of Jeffrey conditioning may be initiated by changes in the probabilities of several propositions. Even when E is intuitively confirmatory, an increase in E's probability may be accompanied by decrease in probability of an even-more-strongly confirmatory proposition F, resulting in a decrease of probability for H overall. Zynda considers modifying the condition to specify that no other positively relevant proposition is decreased in probability; however, he shows by example that even this modified condition excludes some cases in which E intuitively confirms H.

Zynda concludes that "Jeffrey conditioning introduces a new, quite radical kind of contextualism: the evidential value a given proposition has cannot be evaluated independently of what is happening in other regions of logical space" (*op. cit.*, p. 77).

> Judgments of evidential support have to be relativized to particular ways of updating, i.e. to particular epistemic events. Once we have a particular epistemic event in mind...judgments of evidential support make sense.... The quite different project of coming up with an analysis of "e confirms h for S at t" is not in and of itself a very fruitful one for Bayesians to pursue (*op. cit.*, p. 79).

If this argument is correct, we should give up the project of analyzing synchronic confirmation relations altogether. Once we admit Jeffrey conditioning, the evidential value of propositions is contextualized in a way that precludes our meaningfully asking in general whether E confirms H for S at t.

It seems to me, however, that the project of analyzing the synchronic notion of confirmation should not be dismissed this quickly. Recall that, in response to the first difficulty, Zynda modified his condition to exclude cases involving a certain kind of interference—cases where the probability of some other positively relevant proposition F was reduced while E's probability was raised. But the example to show that the modified condition was still unacceptable involved the interfering effects of another evidential proposition—a negatively relevant proposition whose probability was raised. This suggests a further modification, in which the condition would also specify that this second sort of interference not occur. The basic idea here would be that, in Jeffrey conditioning, raising the probability of a confirming proposition E will, ceteris paribus, raise the probability of H. The ceteris paribus condition should naturally preclude both simultaneous reduction of probability for other confirming propositions, and simultaneous increase in probability for disconfirming propositions.

Thus remodified, the condition avoids the difficulties that faced its predecessors. So the fact that Jeffrey conditioning can involve multiple evidential propositions simultaneously does not, after all, render the evidential effect of any one of these propositions contextualized in a way that would prevent us from asking meaningfully whether a

[20] See, for example, Glymour; Eells; Charles S. Chihara "Some Problems for Bayesian Confirmation Theory," *British Journal for the Philosophy of Science*, XXXVIII (1987): 551-60; and Earman, *Bayes or Bust?*

[21] See Maher.

[22] The classical conditionalization principle changes a probability function in response to an evidential proposition's being raised to 1. Jeffrey conditioning changes a probability function in response to initiating changes in the probabilities of one or more propositions. These initiating changes may be in either direction, and may be to nonextreme probability values.

given evidential proposition confirms or disconfirms a given hypothesis (for a given agent at a given time). Indeed, it seems that allowing Jeffrey conditioning leaves us in much the same position with respect to the synchronic notion of confirmation.[23]

VI. A SOLUTION TO THE SYNCHRONIC PROBLEM?

Why does increasing pr(E) drain E of confirmatory power, as measured by the standard account? The standard account measures confirmation by how much increasing pr(E) to 1 would raise pr(H). This partly realizes the traditional confirmation-theoretic idea that E's confirming H does not require E to be believed. But, curiously, it does not fully accord with the idea that our current degree of belief in E is irrelevant to the question of how much E confirms or disconfirms H. Consider a situation in which an agent's beliefs change by her becoming increasingly confident in E. As pr(E) increases, pr(H) increases to take account of the increased confidence in E. Thus, the probability of H on the assumption that E is true, and the probability of H all things considered, become closer. In a sense, E just becomes part of "all things considered." At this point, becoming more certain of E cannot raise the probability of H much more, since there just is not room to become much more confident in E. In such circumstances, the standard measure shows very little evidential support. Yet intuitively, the agent's confidence in E is already an important support for her present confidence in H. This dimension—we might call it the dimension of *actualized support*, as opposed to *potential further support*—is simply ignored by the standard measure. This is the problem of old evidence.

From this perspective, we can also see why the traditional measure suffers from what might be called the *probable-hypotheses problem*. This is a difficulty that is very similar to the old-evidence problem, but much less celebrated.[24] Consider cases where pr(H) gets very high. Again, there is in such cases little room for E (any E) to raise pr(H) further. The traditional measure, taking into account only potential further support, reflects this fact; once pr(H) is close to 1, no evidence will be counted as providing a significant level of support for H. Oddly, once we become quite confident in a hypothesis, we are counted as having no significant evidential basis for our confidence!

It seems crucial, then, that our measure of confirmation go beyond measuring potential further support. The question then arises whether a dimension of actualized support is somehow reflected in the agent's current probability function. Intuitively, the standard measure reflects the amount that pr(H) would rationally increase just as a result of the agent raising her pr(E) to 1. This potential increase in pr(H) might be thought of as resulting from two factors: the degree to which E is linked to H, and the distance E has to travel in attaining probability 1. What we are interested in is the former factor, not the latter.

How might we control for the latter factors' influence? The most obvious way would be to normalize in a way that corrects for it. For example, if pr(E) is already .75, so that it can be raised by only 1/4 of the distance between 0 and 1, we can multiply the potential increase in pr(H) by 4. If pr(E) is already .99, so that it can be raised by only 1/100 of the distance between 0 and 1, we would multiply the potential increase in pr(H) by 100. It turns out that dividing the standard measure by pr($\sim E$) achieves just this result. This suggests using the following as a measure of confirmational support:

$$S(E,H) = [\mathrm{pr}(H/E) - \mathrm{pr}(H)]/\mathrm{pr}(\sim E)$$

Now, one might immediately object to this measure that, when pr(E)=1, it is undefined. But as noted above, there are good reasons to hold that contingent propositions should never be given probability 1. Thus, we may stipulate that the definition—which is, after all, part of an account of ideal reasoning—apply only when $0<\mathrm{pr}(E)<1$. This tack is especially appealing if one is trying to analyze confirmation in terms of standard probabilistic relations among an agent's beliefs, since from this point of view, propositions having probability 1 are indistinguishable from tautologies. This is important because it strongly suggests that on a standard probability-based approach, it would be hopeless to try to measure some interesting kind of evidential support from such propositions.

Interestingly, there is another intuitive route to essentially the same measure. We want to capture the support an agent's confidence in H *already* receives from E (in contrast to the potential further support that might be gotten from raising pr(E)). We might

[23] This impression is reinforced by the observation that the remodified condition will be satisfied whenever the standard positive-relevance condition is satisfied. My hunch to this effect was proved by Zynda (in correspondence; I should also note that he expressed his current reservations about the strong claim criticized above). A related theorem is proved by Carl G. Wagner in "Old Evidence and New Explanation," *Philosophy of Science*, LXIV (1997): 677-91.

[24] This problem—or at least the qualitative version that occurs when pr(H)=1—is briefly celebrated by Kaplan (p. 50), who notes its similarity to the old-evidence problem.

measure this by comparing the agent's present probability for H—which includes that present support—with her probability for H *on the condition that E is not true*: $\mathrm{pr}(H) - \mathrm{pr}(H/{\sim}E)$. Adding this dimension of actualized support to the traditional measure of potential further support yields:

$$S^*(E,H) = \mathrm{pr}(H/E) - \mathrm{pr}(H/{\sim}E)$$

as a measure combining both sorts of support. If conditional probabilities are defined in the standard way, S^* is also undefined when $\mathrm{pr}(E)=1$. And in cases where $0<\mathrm{pr}(E)<1$, measure S^* turns out to be equivalent to measure S.[25]

S^* is briefly considered, and rejected (for reasons that will be examined below) by Earman.[26] His reason for considering it is not either of the intuitive arguments given above. Rather, he points out that conditional probabilities can be defined in a (nonstandard) way which allows $\mathrm{pr}(H/{\sim}E)$ to be defined when $\mathrm{pr}(E)=1$. If we do this, then S^*, unlike the standard positive-relevance measure, can take positive values when $\mathrm{pr}(E)=1$. But the measure has other advantages, in addition to the intuitive naturalness explained above, which strike me as more important. Since the remaining discussion will not turn on interpreting conditional probabilities nonstandardly, I shall treat S^* as an alternate formulation of S, and

[25]

$$S(E, H) = \frac{\mathrm{pr}(H/E) - \mathrm{pr}(H)}{\mathrm{pr}({\sim}E)}$$

Expanding $\mathrm{pr}(H)$, we get:

$$\frac{\mathrm{pr}(H/E) - [\mathrm{pr}(H/E)\mathrm{pr}(E) + \mathrm{pr}(H/{\sim}E)\mathrm{pr}({\sim}E)]}{\mathrm{pr}({\sim}E)}$$

Regrouping the numerator gives:

$$\frac{[\mathrm{pr}(H/E) - \mathrm{pr}(H/E)\mathrm{pr}(E)] - \mathrm{pr}(H/{\sim}E)\mathrm{pr}({\sim}E)}{\mathrm{pr}({\sim}E)}$$

Factoring $\mathrm{pr}(H/E)$ out of the left side of the numerator yields:

$$\frac{\mathrm{pr}(H/E)[1 - \mathrm{pr}(E)] - \mathrm{pr}(H/{\sim}E)\mathrm{pr}({\sim}E)}{\mathrm{pr}({\sim}E)}$$

Since $[1 - \mathrm{pr}(E)] = \mathrm{pr}({\sim}E)$, this may be canceled out, leaving:

$$\mathrm{pr}(H/E) - \mathrm{pr}(H/{\sim}E) = S^*(E, H)$$

[26] *Bayes or Bust?*, pp. 120-21.

shall use whichever formulation is best suited to the purpose at hand.[27]

As Earman notes, S agrees qualitatively with the standard measure in cases involving nonextreme probabilities. (This can be seen most easily by considering the first formulation given above. Since S simply divides the standard measure by a positive number ($\mathrm{pr}({\sim}E)$), positive (negative) confirmation will remain positive (negative).) Given that the standard measure is qualitatively plausible in cases involving nonextreme probabilities, this is an advantage.

More impressive, however, is S's quantitative treatment of old evidence. The most troubling version of the old-evidence problem occurs when $\mathrm{pr}(E)$ is high, but still short of unity. Unlike the standard measure, S allows *significant* confirmation even when $\mathrm{pr}(E)$ is, say, .999999. This is an enormous advantage. It also suggests that S may come closer than the standard measure to satisfying the intuition that the degree to which E confirms H should not depend on how confident we are of E.

In fact, S is, in an intuitively appealing way, insensitive to $\mathrm{pr}(E)$. Suppose an agent has a probability function relative to which the support E gives to H as measured by S (let us call this *S-support*) has some particular value. Let us change the agent's probabilities by Jeffrey conditioning initiated by raising (or lowering) just the probabilities for E and ${\sim}E$. It would be nice if, relative to the agent's new probability function, the S-support E gives to H retained the same value. In fact, this is the case: in general, S-support given by E is stable over Jeffrey conditioning on $\{E,{\sim}E\}$.[28]

As an illustration, consider a rational agent confronted by a standard die he believes to be fair, which has been rolled, but is out of sight. Let O be the proposition that an odd number came up, and L be the proposition that a low number (1-3) came up. For such an

[27] As this paper was under consideration, Jim Joyce sent me a manuscript chapter of his forthcoming book, in which he supports a version of measure S^* as capturing one concept of evidential relevance. Joyce would apply the measure even in cases where $\mathrm{pr}(E)=1$, by defining it via Reyni-Popper measures, which work much like standard conditional probabilities, but define $\mathrm{pr}(H/{\sim}E)$ even in cases where $\mathrm{pr}({\sim}E)=0$. Joyce interprets such measures as describing the epistemic state of someone who is fully certain that E is true, but still has opinions as to the probability of H if she is wrong about E. Joyce's discussion is highly recommended. While it has some parallels with the present discussion, it has interesting differences and reaches different conclusions; for example, he sees the standard confirmation measure as capturing a different, but no less legitimate, concept of evidential relevance. See Joyce's *The Foundations of Causal Decision Theory* (New York: Cambridge, 1999), chapter 6.

[28] This is easy to see in the S^* formulation. The conditional probabilities by which the measure is defined are unchanged in Jeffrey conditioning, since Jeffrey conditioning involves rigidity of all probabilities conditional on the members of the relevant partition.

agent, $\mathrm{pr}(O)=1/2$ and $\mathrm{pr}(L)=1/2$. Using the second formulation given above, $S(L,O) = \mathrm{pr}(O/L) - \mathrm{pr}(O/{\sim}L) = 2/3 - 1/3 = 1/3$: a modest degree of positive support.

Now suppose the agent raises the probability of L to .99 by Jeffrey conditioning in which L and ${\sim}L$ are the only propositions directly affected by experience. Her new probability PR for O will, of course, be significantly higher:

$$\mathrm{PR}(O) = .99\mathrm{pr}(O/L) + .01\mathrm{pr}(O/{\sim}L) = .66 + .00333... = .66333...$$

This higher probability for O intuitively incorporates the agent's increased confidence in L. It also renders support as standardly measured almost nonexistent: $\mathrm{PR}(O/L) - \mathrm{PR}(O) = .00333....$ This is a prime example of the synchronic old-evidence problem. The low support figure represents the potential for becoming more certain of O on the basis of becoming even more certain of L, but leaves out the extent to which $\mathrm{PR}(O)$ already incorporates support from confidence in L.

By contrast, S-support is unchanged at 1/3. Thinking in terms of the first formulation, it is easy to see how this figure represents the standard measure of support normalized by being multiplied by 100, since the agent's move to full belief in L would move $\mathrm{PR}(L)$ 1/100 of the way from 0 to 1. Intuitively, it seems that S counts both potential and actualized support. Cases like this suggest that S offers a possible solution to the synchronic problem of old evidence.[29]

VII. THE OLD-EVIDENCE PROBLEM RETURNS?

As noted above, Earman quickly rejects S^* as a suitable measure of confirmation. His reason is that, in cases where H entails E and

[29] It may be of interest to see how this sort of situation is handled by the likelihood-ratio-based model mentioned above. The model would measure confirmation in the dice case by contrasting the likelihoods $\mathrm{pr}(L/O)$ and $\mathrm{pr}(L/{\sim}O)$. In the initial situation, these quantities are 2/3 and 1/3 respectively, so the likelihood ratio is 2. But after the probability of L is raised to .99, the likelihoods take on new values. In particular (using values from the text):

$$\mathrm{PR}(L/O) = \frac{\mathrm{PR}(L)\cdot\mathrm{PR}(O/L)}{\mathrm{PR}(O)} = \frac{.99\cdot(2/3)}{.66333...} = \frac{.66}{.66333...} \approx .99$$

$$\mathrm{PR}(L/{\sim}O) = \frac{\mathrm{PR}(L)\cdot\mathrm{PR}({\sim}O/L)}{\mathrm{PR}({\sim}O)} = \frac{.99\cdot(1/3)}{.33666...} = \frac{.33}{.33666...} \approx .98$$

Thus the likelihood ratio shrinks to about $.99/.98 \approx 1.01$.

$\mathrm{pr}(E)=1$, $S^*(E,H)$ is just equal to $\mathrm{pr}(H)$. Thus, for any hypothesis H, all evidence which is both (1) entailed by H and (2) believed with certainty is counted as confirming H to the same degree.

Now, this problem as posed by Earman might not seem to concern us if we intend our measure to be restricted to cases where $\mathrm{pr}(E)<1$. But there is a closely related phenomenon that affects even these cases. It derives in part from the fact that, in cases where $H \vdash E$, $S(E,H) = \mathrm{pr}(H/E)$ (this is because $\mathrm{pr}(H/{\sim}E) = 0$). This fact alone does not make all entailed evidence equal, however. For example, consider our agent who has ordinary expectations about a fair unseen die, and who thus has the standard simple and conditional probabilities for the following propositions:

$E2$: the die shows a multiple of two.
$E3$: the die shows a multiple of three.
$H6$: the die shows six.

Both $E2$ and $E3$ are entailed by $H6$. In this situation, both should confirm $H6$, but $E3$ should confirm it more strongly. And measure S gives just this result: $S(E2,H6) = \mathrm{pr}(H6/E2) = 1/3$, while $S(E3,H6) = \mathrm{pr}(H6/E3) = 1/2$.

So far, we have not seen the difficulty. But, as we know from the old-evidence problem, as $\mathrm{pr}(E)$ approaches 1, $\mathrm{pr}(H/E)$ approaches $\mathrm{pr}(H)$. Thus, if the above agent's probabilities for both $E2$ and $E3$ are raised to nearly 1, the probability of $H6$ conditional on each of these propositions will be very close to the unconditional probability of $H6$, and hence the support the two propositions give $H6$ will, according to S, be very nearly the same. This is a disturbing result, for intuitively, $E3$ should give more confirmation than $E2$, even when both are highly probable. Does this not show that S encounters the old-evidence problem after all, at least for cases where $H \vdash E$?

Closer inspection reveals, I think, that the answer to this question is "no." This is not the problem of old evidence. To see this, let us begin by considering what might seem an inconsistency between the claims made above about S.

Immediately above, we saw that, in a situation involving ordinary beliefs about a hidden die, S assigns $E2$ and $E3$ robustly different degrees of support for $H6$, while if $E2$ and $E3$ both become very highly probable, S will have to assign them very nearly the same degree of support for $H6$. In the previous section, however, we saw that S is, in a very natural sense, insensitive to $\mathrm{pr}(E)$. If this is true, how could raising $\mathrm{pr}(E2)$ and $\mathrm{pr}(E3)$ erase the difference in the S-support they give $H6$?

The first thing to notice is that the previous section's claim does hold in the present example. If we begin with ordinary beliefs about a hidden die, and increase belief in, for example, $E2$ by Jeffrey updating on just $E2$ and its negation, $S(E2,H6)$ does not change, even when $\mathrm{pr}(E2)$ gets very high. The probabilities in such a situation will look like this (using PR for the post-updating beliefs, and pr for the initial beliefs):

$$\mathrm{PR}(E2) = .99$$
$$S(E2,H6) = \mathrm{PR}(H6/E2) = \mathrm{pr}(H6/E2) = 1/3$$

This shows that $S(E2,H6)$ remains unchanged. But other things have changed. For one thing, $\mathrm{PR}(H6)$ now reflects the positive evidence from $E2$:

$$\mathrm{PR}(H6) = .99\mathrm{pr}(H6/E2)+.01\mathrm{pr}(H6/{\sim}E2) = .99(1/3)+0 = .33$$

More interestingly, some probabilities conditional on $E3$ have changed. The new probability for $E3$ itself is still 1/3, since $E2$ is probabilistically irrelevant to $E3$. But:

$$\mathrm{PR}(H6/E3) = \mathrm{PR}(H6 \ \&\ E3)/\mathrm{PR}(E3) = \mathrm{PR}(H6)/\mathrm{PR}(E3) = .33/(1/3) = .99$$

Because of this, the support $E3$ gives to $H6$, according to measure S, is now close to maximal:

$$S(E3,H6) = \mathrm{PR}(H6/E3) = .99$$

So, while raising the probability of $E2$ did not change *its* confirmatory power for $H6$, it dramatically increased the confirmatory power of $E3$. Raising the probability of one evidential proposition can thus affect confirmation relations involving other propositions. So the fact that $S(E, H)$ is insensitive to simply raising the probability of E is quite consistent with the fact that raising the probability of two or more evidential propositions will disturb the S-confirmation relations involving all of them.

This example also shows that the disturbing fact—that raising the probabilities of both $E2$ and $E3$ would erase the differences in confirmatory power between them—is not simply an instance of the old-evidence problem. For either one of these same propositions can be believed with near certainty without any effect on its confirmatory power. Still, while measure S thus seems to *avoid* the old-evidence problem, it would be premature to conclude that it *solves* the problem. Presumably, a solution to the problem would require a measure which not only avoided it, but which was otherwise reasonable. How unreasonable are the verdicts S delivers in the disturbing cases?

VIII. A DIFFERENT PROBLEM FOR BAYESIANISM

Let us begin by looking more carefully at the problematic case mentioned above. Suppose that an agent starts with ordinary beliefs about a hidden die, and is told first by one highly trusted source that it shows a multiple of two ($E2$), and then by another that it shows a multiple of three ($E3$). Intuitively, $E3$ would seem more important than $E2$ in supporting her now-high confidence that ($H6$) the die is showing a six. But measure S must now give them almost equal confirmatory power. In particular, the confirmatory power of both $E2$ and $E3$ is nearly maximal in this case.[30] Raising the probability of $E2$ raised the confirmatory power of $E3$, and vice versa, to the extent that the difference between them was nearly erased.

One important question to be asked about this example is about exactly where the counterintuitiveness lies. The fact that $E2$ and $E3$ have nearly equal confirmatory power derives from another fact: that each of them has confirmatory power close to the maximum of 1. But this seems to be a problematic fact about each one of the evidential propositions, independent of questions about *differences* in confirmatory power. (If it was intuitively reasonable that raising $\mathrm{pr}(E2)$ would make $E3$ nearly maximally confirmatory, and vice versa, then presumably it would be reasonable for them to have approximately equal confirmatory power in that situation.) If this is right, then the real root of the problem should reveal itself midway through the example story, when the agent comes to believe strongly in $E2$, rendering $E3$ almost maximally confirmatory. What, intuitively, is going on at this point?

Recall that the way learning $E2$ raises the confirmatory power of $E3$ so dramatically is something like this: once $\mathrm{pr}(E2)$ becomes close to 1, $\mathrm{pr}(H6/E3)$ gets close to 1. Since $S(E3,H6)$ is in this case equivalent to $\mathrm{pr}(H6/E3)$, it also rises to nearly 1. Or, more informally, what happens is that when it is almost certain that a multiple of two is showing, the information that a multiple of three is showing is virtually equivalent to the information that a six is showing. Our confirmation measure reflects the strength of this connection.

Why should this sharp increase in confirmatory power be counterintuitive? I think it is useful to compare this case to another. Consider a second agent, who starts off with beliefs appropriate to a fair die having only three (curved) sides, bearing a two, a four, and a six, respectively. For such an agent, it would intuitively be entirely reasonable to take $E3$ as maximally confirmatory. After all, with these

[30] Consider $E2$: $\mathrm{pr}(H6/E2)$ is almost 1, since it is higher than $\mathrm{pr}(H6)$, which itself is near 1. And $\mathrm{pr}(H6/{\sim}E2)$ is 0, because $H6 \vdash E2$. Thus $S(E2,H6)$ is almost 1.

background beliefs, $E3$ is virtually equivalent to $H6$. But now let us ask what the difference is between our two agents. From the point of view of their probabilities, the answer seems to be "not much!" For both, pr($E2$) would be nearly 1, pr($E3$) and pr($H6$) close to 1/3, and pr($H6/E3$) nearly 1. For both, learning $E3$ would virtually guarantee $H6$. Why, then, is there such a difference in our intuitions about confirmatory power in these two cases?

The answer, it seems to me, is related to the following intuition: in the case with the ordinary die, if $E3$ is learned, it will raise pr($H6$) to near 1, but it will be doing only part of the work of raising it from its baseline level (1/6), the rest being done by $E2$. In the three-sided die case, if $E3$ is learned, it will be doing all of the work in raising the probability of $H6$ from its different baseline level (1/3) to near certainty.

When we think about the evidential power of $E3$ in the case with the ordinary die, we seem to assess it against $H6$'s intuitive base-line level. We intuitively take the elimination of the odd-numbered possibilities as something that was accomplished by another piece of evidence. When we think about the three-sided die case, by contrast, we take the absence of odd-numbered possibilities as being simply part of the given background conditions for our reasoning. This suggests that the intuitive difference between the cases depends on a factor to which our confirmation measure is insensitive: the distinction between specific evidence other than E, and background assumptions.

This point constitutes a further departure from the worry expressed by Earman. He rejected a version of measure S because it equalized confirmatory power for all evidence that was both entailed by the hypothesis and had probability 1. We have already seen that essentially the same difficulty occurs at evidential probabilities below 1. Moreover, in one such case, it seems that unintuitive equalization per se is not the basic problem—rather, it is a symptom of unintuitiveness in measure S's verdicts on individual items of evidence. Now, if the essential problem is not about equalizing differences in these cases, but rather about a very general feature of confirmation—the background/evidence distinction—then we might suspect that the problem should occur even in situations where H does not entail E.

It seems to me that the fundamental problem is, in fact, quite independent of H entailing E. To see this, consider the following variant of our deer example: the agent is wondering whether deer live in a certain wood that he has not yet explored. Let H be 'Deer live in this wood', D be 'There are deer droppings at location x', and A be 'There is a shed deer antler at location y', where x and y are some narrowly circumscribed locations in the wood. This case is in a sense diametrically opposite from the dice case, in that the evidential propositions are not even "close to" being entailed by H. (Since D and A specify particular locations for the droppings and antler, their probabilities even conditional on H are not very high.)

Now suppose the agent starts with pr(H)=.5, pr(D)=.001, and pr(A)=.0001. Both D and A will confirm H fairly strongly—at about .5.[31] Suppose, however, that the agent then finds what are almost certainly deer droppings at location x, sending the probabilities for D, and hence H, very high. This does not, of course, change D's level of S-support for H. This is good news. In fact, it not only shows, like the previous example, that S allows old evidence to confirm hypotheses strongly. It also shows that S allows instances where E strongly confirms H even when pr(H) is very high. Thus, S does not seem to suffer from what we called above the problem of probable hypotheses—the inability of the standard measure to assign significant degrees of confirmation to highly probable hypotheses.

Unfortunately, not all the news is good. In particular, although S-support from D is unchanged, S-support from A suddenly becomes very low. This is because H is now very likely irrespective of A, so pr(H/A) and pr($H/\sim A$) are nearly the same. This is quite unintuitive; A seems just as good a sign of deer as it was before D became highly probable. Moreover, once the agent finds the antler and A becomes very probable, even $S(D,H)$ becomes very low. Here, we have an even more counterintuitive case: the agent has quite reasonably come to have high confidence in H, based on learning A and D. But according to measure S, neither A nor D now gives significant support to H (though, oddly, their disjunction does)![32]

This seems, at first blush, to be much like the original problem of probable hypotheses. But it is certainly not precisely the same problem; we have seen that S does not automatically preclude strong con-

[31] In this situation, the probabilities for H conditional on $\sim D$ and on $\sim A$ will be close to the unconditional probability of H—around .5—conforming with the intuition that the fact that there were not, for example, droppings *at location x* would have little effect on the probability of deer living in the wood. But the probabilities for H conditional on A and on D will be close to 1. Subtracting the former conditional probabilities from the latter thus leaves S-support at about .5 in each case.

[32] It is worth noting that this problem is different than that described above for the historical approach. There, after both pieces of evidence were discovered, their evidential weights depended on an irrelevant factor: the order of discovery. The problem here is different, though no less real.

firmation of highly probable hypotheses. What, intuitively, is going wrong?

Again, I think that it will be helpful to look at the example halfway through the story. Let us examine the first counterintuitive result—that the confirmatory power of the antler is drained when discovery of the droppings raises the agent's probability for deer. At this point, the low value measure S assigns to A reflects the fact that, *given the agent's probabilities*, A's truth or falsity is not highly relevant to H. But if A is not highly relevant to H, given the agent's probabilities, why do we have the intuition that A is highly confirmatory? It seems to me that, when we think about the question of A's evidential bearing on H, we are impressed by the fact that, given the agent's general background beliefs about the relations between deer, antlers, and droppings, we can see that, *putting aside the agent's specific evidence D, A* would be highly relevant to H. The intuitive idea—admittedly a vague one—is that the agent's degree of confidence in A would make a big difference to his confidence in H ceteris paribus, where the ceteris paribus condition would hold the agent's general background beliefs constant, but abstract away from the agent's beliefs that are being thought of as his specific evidence about H. The question seems to be something like 'How important would belief in A be, if it were the agent's only evidence about H'?[33]

[33] Maher's account of confirmation involves various contextual factors, among which is a specification of a set of evidential propositions "whose relevance to H is under discussion or otherwise salient in the context" (*op. cit.*, p. 167). But Maher's approach differs markedly from the present one in a number of ways. First, his account is purely qualitative (Maher is unpersuaded by arguments that the standard account of qualitative confirmation is adequate). Thus it makes no attempt to address the quantitative issues with which we have been concerned. Also, Maher's account makes use of other factors which go beyond an agent's probability distribution.

Maher's analysis relativizes confirmation to (contextually determined) "background evidence," which is taken to be so complete that evidence E confirms H relative to that background only if E is positively relevant to H for *every* probability function that could be rational for an agent whose total evidence consisted of that background. The assumption that the relevant background can be contextually determined in this way makes Maher's account objective, in the sense that it is not relativized to an agent's probabilities at all. But this assumption would be rejected by those who see confirmation as depending on past learning which is encoded in an agent's degrees of belief in a way that does not allow one to isolate out the evidence on which those degrees of belief were based. Maher's account also depends on an unanalyzed determination of whether some propositions are "based on an inference from" others. To my mind, this last aspect of the account especially raises troubling questions, since it seems to be presupposing a notion that is fairly close to (and may itself depend on) the notion being analyzed. But more detailed discussion of Maher's account must await another occasion.

To test this diagnosis, it would be natural to construct a situation which is similar to the deer case, but in which the probability of H is high based on background alone. In such a situation, it should seem more reasonable to say that A is not strong evidence for H. But it is a bit hard to do this, for the following reason: what counts as background in a given context will depend on the hypothesis under discussion. If we are discussing H, then what intuitively counts as evidence will naturally include whatever supports the agent's high degree of belief in H. But this makes it hard to think of a case where H is probable on the basis of "background" beliefs alone. The following is, however, an attempt in this direction:

Let H be 'There are at least five trees in the world', and A be 'There are at least four trees in Jocko's back yard'. Suppose first that a person has ordinary beliefs about trees, and thus a very high probability for H, but does not know how many trees are in Jocko's yard. In such a situation, I think that it would be unnatural to say that A would, for that agent, be strong evidence for H. That is because H is so probable anyway, given the agent's background beliefs, that the agent's lack of confidence in A makes little difference. (Here measure S gives the intuitively reasonable verdict, since $pr(H/E)$ is essentially the same as $pr(H/{\sim}E)$.) Were the agent to become highly confident of A, it would still seem unnatural to say that A was good evidence of H for her (and S would still agree with intuition). My suggestion is that our intuitions in this case differ from those in the deer case because H is so probable, given the relevant "background" beliefs, that we do not see confidence in A as making a significant difference to confidence in H ceteris paribus. Measure S gets this one right (and the deer example wrong) because it essentially treats everything other than E as "background."

In sum, then, it turns out that measure S does possess a significant advantage over the traditional measure: it avoids the old-evidence problem (and its first cousin, the probable-hypotheses problem). S cannot be seen as an accurate model of confirmation, however. The reason does not ultimately stem from the particular features with which Earman was concerned, but rather stems from the fact that S is insensitive to an aspect of the way we think about confirmation: the distinction between specific evidence and background assumptions. The inability to account for this feature, of course, is not peculiar to S. It will affect any measure that tries to analyze confirmation based purely on an agent's degrees of belief.[34]

IX. CONCLUSION

Formal models of philosophically interesting concepts are tested largely by seeing how well they match intuitive judgments in particular cases. Persistent failures to find a formal model that matches our intuitive judgments may mean that we have yet to find the technical notion that corresponds to our concept. But they may also indicate that there is some feature of the concept itself which resists formal capture.

In some cases, this can simply be due to indeterminateness. Perhaps the controversy between difference- and ratio-based positive relevance models of quantitative confirmation reflects a natural indeterminateness in the basic notion of "how much" one thing supports another. This sort of resistance to formal capture should not, of course, be seen as undermining the philosophical interest of the formal models. If the models can show us two different ways that an indeterminate concept can be made precise, then they have taught us something.

The problems presented by old evidence (and by probable hypotheses) are not, however, of this sort. Both problems threaten to undermine the usefulness of the traditional model of quantitative confirmation (as well as its ratio-based variants). And old evidence does the same for confirmation models based on likelihood ratios. In this respect, the fact that measure S avoids those problems constitutes an important advantage.

But even measure S, as we have seen, does not succeed in matching our intuitive quantitative confirmation judgments. And the reasons for the mismatches suggest that no purely probabilistic account could match those judgments. The question then arises whether the fact that S avoids even an important problem like the old-evidence problem can have any philosophical importance.

It seems to me that it can. Even models that fail to capture some important features of ordinary conceptions may be quite useful in helping us understand those conceptions, and the phenomena to which those conceptions answer; getting the cases right is not the sole point of a philosophical model. To the extent that a formal model helps one to isolate, identify, and understand those aspects of confirmation which resist formal capture, the model renders important philosophical service.

S is best seen not as a correct account of confirmation, but as a useful tool for understanding confirmation. It models one of the central aspects of confirmation—the one involving E making H more probable. Moreover, it does this while avoiding some problems that limit the usefulness of the traditional probabilistic measure (and the other popular measures we have considered). Its avoidance of these problems has helped us see, by examining the very cases where S gives unintuitive verdicts, other aspects of our conception. It has long been clear that the traditional probabilistic account of confirmation would fail to capture the epistemic asymmetry of our ordinary notion. It now seems that probabilistic accounts will also miss our ordinary notion's dependence on the distinction between background beliefs and specific evidence.

I would not claim that these are the only aspects of the intuitive conception that the measure fails to match. But we should not hastily take further mismatches as vitiating the philosophical interest of probabilistic models of quantitative confirmation. For in examining the cases in which our formal model gives unintuitive results, and putting our fingers on the reasons for the mismatches, we may yet reveal to ourselves more clearly the contours of confirmation.

DAVID CHRISTENSEN

University of Vermont

[34] The examples considered in the text make clear that we could not avoid this problem even if we enriched our representation of a probability function in a way that would allow us to recover a "pure" initial probability function and each bit of evidence that had combined with that initial function to produce the agent's present function (see Brian Skyrms, "Three Ways to Give a Probability Assignment a Memory," in Earman, ed., *Testing Scientific Theories*, pp. 157-61). For the background beliefs which underlie the probabilistic connections between evidence and hypothesis—for example, that deer shed antlers and drop droppings—are themselves based on evidence. Using the "pure" function would dissolve these connections. On the other hand, using a function that resulted from combining the initial function with all the evidence except the particular item in question would result in seeing no significant confirmation in cases like the one involving both antlers and droppings.