Journal of Philosophy, Inc.

Calibration and the Epistemological Role of Bayesian Conditionalization
Author(s): Mark Lange
Source: *The Journal of Philosophy*, Vol. 96, No. 6 (Jun., 1999), pp. 294-324
Published by: Journal of Philosophy, Inc.
Stable URL: http://www.jstor.org/stable/2564680
Accessed: 17/03/2010 00:28

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

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=jphil.

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

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

Journal of Philosophy, Inc. is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Philosophy.*

http://www.jstor.org

CALIBRATION AND THE EPISTEMOLOGICAL ROLE OF BAYESIAN CONDITIONALIZATION *

Suppose that a scientist acquires some new evidence; she learns that *e* obtains. She must update her confidence in hypothesis *h* accordingly. *Bayesian conditionalization* (BC) holds that her new confidence in *h* should equal the confidence that she had (prior to discovering *e*) that *h* obtains in the event that *e* obtains. More precisely, if $\mathrm{pr}_{\mathrm{old}}(\cdot)$ is our probability distribution before discovering *e*, and $\mathrm{pr}_{\mathrm{new}}(\cdot)$ is our probability distribution after discovering *e*, then BC says that:

> *If* $0 < \mathrm{pr}_{\mathrm{old}}(h) < 1$ (so that before discovering *e*, we regarded *h* as a live hypothesis, subject to confirmation and disconfirmation)
> $\mathrm{pr}_{\mathrm{old}}(e) < 1$ (so that *e* was not believed already)
> *e* is the strongest claim whose truth we take ourselves to have discov ered
> and $\mathrm{pr}_{\mathrm{old}}(e) > 0$ (so that $\mathrm{pr}_{\mathrm{old}}(h|e) = \mathrm{pr}_{\mathrm{old}}(h \,\&\, e)/\mathrm{pr}_{\mathrm{old}}(e)$)
> *then* $\mathrm{pr}_{\mathrm{new}}(h) = \mathrm{pr}_{\mathrm{old}}(h|e)$.

Coupled with Bayes's theorem–$\mathrm{pr}_{\mathrm{old}}(h|e) = \mathrm{pr}_{\mathrm{old}}(h)\ \mathrm{pr}_{\mathrm{old}}(e|h)/\mathrm{pr}_{\mathrm{old}}(e)$– BC seems able to account elegantly for many important features of theory confirmation in science, such as the value of diverse evidence, the kernel of correctness in hypothetico-deductivism, the roles of auxiliary hypotheses and rival hypotheses and background beliefs, and the failure in certain cases of hypotheses to be confirmed by their positive instances.[1]

In section I, I review certain considerations often mentioned as posing difficulties for BC. Curiously, several of these purported challenges involve cases that, strictly speaking, fall outside BC's jurisdiction. Why, then, are these considerations widely perceived as making trouble for BC? At least partly because (I suggest) we expect BC to capture not merely our intuitions regarding how we should *revise* our opinions (that is, update our degrees of confidence in various claims) upon discovering new evidence, but also our intuitions regarding whether *e counts as evidence for h*; we tend to regard BC as purporting to explain why vari-

* My thanks to Jeff Barrett, Jim Bogen, Mitch Green, Alan Hajek, Michael Otsuka, Ric Otte, Jim Woodward, and Lyle Zynda for discussions of early drafts, and to Bill Talbott for more recent discussion.

[1] Many authors have argued that BC can account for these features. For discussion, see John Earman, *Bayes or Bust?* (Cambridge: MIT, 1992), chapter 3 ("Success Stories"). A careful examination of these arguments and their potential defects is beyond the scope of this paper.

0022-362X/99/9606/294–324 © 1999 The Journal of Philosophy, Inc.

ous factors make *e* stronger or weaker evidence for or against *h*. But (as we shall see) the notion of *e*'s being 'evidence for *h*' is not restricted to cases where we discover *e*, $0 < \mathrm{pr}_{\mathrm{old}}(e) < 1$, $0 < \mathrm{pr}_{\mathrm{old}}(h) < 1$, and we update our confidence in *h* accordingly. So even if BC governs every such updating, it does not thereby automatically account for every case in which *e* constitutes evidence for (or against) *h*.

In section II, I distinguish two roles that BC might play in an account of confirmation: as governing how we should *update* our opinions upon discovering *new* evidence, and as governing the steps in the *arguments* by which our current opinions are to be justified by the evidence that we have *already* assembled. The former is a diachronic matter; the latter concerns the arguments that we give at a particular moment for our opinions then. Although this is an elementary distinction, I have not seen it drawn elsewhere. This distinction is useful, I argue, because if BC governs justificatory arguments, then it is in the right position to cash out what it is for *e* to be evidence for (or against) *h*. In this role, BC could account for the various features of confirmation that it seems able to explicate so nicely, and yet avoid the problems that result from trying to explicate '*e* is evidence for *h*' in terms of the difference that *e*'s discovery makes, made, or would have made to our confidence in *h*.[2]

In section III, I consider *why* it would be wrong for a justificatory argument to violate BC–what sort of logical error would thereby be made. Diachronic Dutch Book arguments have sometimes been thought to explicate why it would be wrong for us to violate BC in updating our opinions. But these arguments have encountered formidable obstacles. Indeed, for reasons I discuss in section I, I think that it is *rationally permissible* for an *update* to violate BC. I use the notion of *calibration* to elaborate the sense in which it would be wrong for a *justificatory argument* to violate BC. I demonstrate that a scientist's opinions at two steps of a justificatory argument can all be calibrated (in the relevant sense) only if her transition from one step to the next accords with BC. This calibration argument, I contend, avoids difficulties encountered by diachronic Dutch Book arguments.

In applying BC to justificatory arguments rather than to diachronic updates, I aim to find a role for BC to play where it could account for various important features of confirmation.

[2] I do not contend that all serious obstacles to reconstructing scientific practice in accord with BC can be avoided in this way. In particular, logical omniscience (which I assume throughout) and the problem of the priors require separate treatment. Also, I presume throughout that upon making an observation, we accept a claim whose content expresses our new evidence, and that any evidence presented in a justificatory argument consists of a proposition whose truth has been ascertained.

I

I shall begin by briefly reviewing four alleged problems for BC and trying to understand just why they are supposed to be problems.

First, consider the *problem of old evidence.* Suppose that h was formulated only *after* e was discovered. Since we already believe e by the time we formulate h, our $\mathrm{pr}_{\mathrm{old}}(e)$ at that time is 1, so $\mathrm{pr}_{\mathrm{old}}(e|h) = 1$, so (by Bayes's theorem) $\mathrm{pr}_{\mathrm{old}}(h|e) = \mathrm{pr}_{\mathrm{old}}(h)$, and so (by BC) $\mathrm{pr}_{\mathrm{new}}(h) = \mathrm{pr}_{\mathrm{old}}(h)$. Yet we may well regard e as evidence for (or against) h.

So the problem of old evidence is typically posed.[3] But BC applies only when, having already formulated h, we discover e and must update our confidence in h accordingly; recall that BC is specifically restricted to cases where $\mathrm{pr}_{\mathrm{old}}(e) < 1$. When e is old evidence, there is no moment at which e's discovery elicits a shift in our confidence in h, since there is no $\mathrm{pr}_{\mathrm{old}}(h)$, prior to discovering e, from which to shift. Old evidence therefore falls outside of BC's official scope. This makes it unclear why old evidence is commonly perceived as posing a problem for BC.

I suggest that one reason why old evidence is taken to be a problem for BC is because BC is intended to account for the various considerations that may make e evidence for h. Old evidence is a problem for this project because intuitively, old evidence can be *evidence for h,* even though old evidence cannot provoke our *updating* our confidence in h, and so BC apparently tells us nothing about what makes some old evidence e stronger or weaker evidence for or against h. In other words, although BC applies only to updatings, Bayesians purport to be giving an analysis of what it is for evidence to "count for or against a theory, or be neutral towards it," as Colin Howson and Peter Urbach[4] put it; it is quite common to treat 'e is evidence for h' as one way of expressing the target of confirmation theory.[5] But no account–Bayesian or otherwise–that applies only to updatings can *straightforwardly* explain when e counts as evidence for h, since not all evidence for h provokes, upon its discovery, an increase in our confidence in h–as when e is discovered before we have even formulated h. Thus, old evidence is a problem for BC.

One way to make an account of updating explain when old evidence e counts as evidence for h, even though our confidence in h is not *actually* updated upon our discovering e, is to apply the account of updat-

[3] For a good review of the literature, see Earman, pp. 119-36.

[4] *Scientific Reasoning: The Bayesian Approach* (Chicago: Open Court, 1993, 2nd edition.), p. 117.

[5] See, for instance, Paul Horwich, *Probability and Evidence* (New York: Cambridge, 1982), p. 51; and Mark Kaplan, *Decision Theory as Philosophy* (New York: Cambridge, 1996), p. 48.

ing to a *counterfactual* updating: how we should have shifted our confidence in h upon discovering e, *had* we formulated h before discovering e. This appeal to counterfactual updatings has been extensively criticized, though not all find these criticisms persuasive.[6] I cannot now give this proposal the scrutiny it deserves, but I need to mention one of its chief difficulties. What our $\mathrm{pr}_{\mathrm{old}}(h)$ would have been, had we formulated h sometime before discovering e, depends upon the background beliefs that we held before discovering e regarding auxiliary hypotheses and alternatives to h. These background beliefs may be quite foreign to us when we judge, long after e was discovered, whether e is evidence for h. Consequently, such historical considerations seem out of place–intuitively irrelevant to e's bearing on h.

Of course, we might avoid this problem by applying BC not to the probability distribution that we would have held prior to discovering e, had h already been formulated by then, but to some other probability distribution, which is not picked out by some exercise in counterfactual history. Presumably, by proper selection of this reference probability distribution, BC could be made to yield verdicts that agree with our intuitions. But we might well wonder why e's status now as evidence for h depends upon the way that we should have updated our opinions upon discovering e, were we beginning from a certain probability distribution that is neither our actual distribution at some past moment nor even the distribution that we would have held then, had h been formulated by then. I shall later suggest that the reference probability distribution is relevant because it appears in the argument by which we justify our current degree of confidence in h. Therefore, whether e is evidence for h depends not on some hypothetical updating, but on our current justificatory arguments.

If we cash out 'e is evidence for h' in terms of some actual or counterfactual updating, depending upon whether e is new or old evidence, then we run the risk of construing 'e is evidence for h' very differently in these two cases. Intuitively, however, the temporal order should not enter into it. I shall ultimately give 'e is evidence for h' a uniform treatment, whatever the temporal order, by elaborating it in terms of e's role in an argument justifying our current degree of confidence in h. Actually, even when e was discovered *after* h was formulated but nevertheless some time ago, we should not construe 'e is evidence for h' in terms of the actual updating that was then provoked by e's discovery.[7] Whether e

[6] For critique, see Earman, pp. 123, 134, and the references cited there. For defense, see, for example, Howson and Urbach, pp. 403-08.

[7] Ellery Eells terms this e 'old new evidence' in his "Problems of Old Evidence," *Pacific Philosophical Quarterly*, LXVI (1985): 283-302.

now counts as evidence for h does not depend on whether *at the time it was discovered*, e provoked an updating that raised our confidence in h. That past updating reflected our opinions *then* regarding various auxiliary and rival hypotheses, whereas it is our *current* opinions that influence whether e *now* counts as evidence for h. (Of course, whether e now counts as evidence for h cannot be fixed by whether, right now, $\mathrm{pr}(h|e) > \mathrm{pr}(h)$, since e was discovered some time ago, so these two probabilities are equal.) Again, I shall suggest that 'e is evidence for h' should be understood in terms of e's role in arguments by which we now justify our current degree of confidence in h.

The second of the four alleged problems for BC that I wish to discuss is that, if h is already believed (that is, $\mathrm{pr}_{\mathrm{old}}(h) = 1$), then discovery of some new evidence e cannot confirm h (that is, cannot raise our confidence in h) under BC (or any other updating rule that respects the axioms of probability). Likewise, e cannot disconfirm h if $\neg h$ is already believed (that is, $\mathrm{pr}_{\mathrm{old}}(h) = 0$). Again, although this is reputed[8] to be a problem for BC, it is initially unclear just why. Like old evidence, this case falls outside BC's jurisdiction, since BC requires $0 < \mathrm{pr}_{\mathrm{old}}(h) < 1$. One reason it poses a problem for BC, I suggest, is again that we expect BC to provide an analysis of 'e is evidence for h', and this project is threatened when e intuitively supports (or is evidence against) h and yet e's discovery fails to result in our updating our opinion of h. Once again, I shall shortly suggest that the fault lies not with BC, but with using it to cash out 'e is evidence for h' in terms of updating rather than in terms of an argument justifying our current opinion of h.

The third alleged problem for BC is a sort of dual to the second: if h is already believed (that is, $\mathrm{pr}_{\mathrm{old}}(h) = 1$), then according to BC, e's discovery cannot disconfirm h—cannot lower our confidence in h—and if $\neg h$ is already believed (that is, $\mathrm{pr}_{\mathrm{old}}(h) = 0$), then according to BC, e cannot confirm h. But this is inconsistent with scientific practice; new evidence sometimes leads scientists to take seriously a theory that they had rejected or to challenge a theory that they had accepted.

Again, however, it is not initially apparent why these considerations make any trouble for BC. As Isaac Levi[9] has noted, BC purports to cover only shifts of opinion upon discovering some *e that is consistent with what we already fully believe.* If $\mathrm{pr}_{\mathrm{old}}(e) = 0$, as when h logically entails $\neg e$ and $\mathrm{pr}_{\mathrm{old}}(h) = 1$, then our rejection of h upon discovering e is entirely consistent with BC, just outside BC's scope. For that matter, BC requires $0 < \mathrm{pr}_{\mathrm{old}}(h) < 1$, so any rise of suspicions re-

[8] See, for instance, Kaplan, p. 50.

[9] *The Enterprise of Knowledge* (Cambridge: MIT, 1980), p. 82.

garding previously accepted hypotheses or rehabilitation of previously rejected hypotheses goes beyond BC's jurisdiction. Such shifts of opinion are consistent with BC.

One reason, I think, that such shifts of opinion are often perceived to pose a problem for BC is again that BC is expected to explain why e is strong or weak evidence for or against h. Intuitively, the factors relevant to e's evidential power are the same whether h is already fully believed, already rejected, or being entertained as a live hypothesis. But an account of 'e is evidence for h' which uses BC to cover the cases where h is a live hypothesis, and some other principles for the other cases, threatens to treat these cases in a counterintuitively nonuniform way. We might well be prepared to grant that along with BC to cover shifts of opinion upon consistent extension of our full beliefs, there are further rules covering *updates* that involve rehabilitating formerly rejected hypotheses or challenging formerly accepted hypotheses.[10] These shifts of opinion are intuitively quite different from "business as usual." But we may be much less willing to grant that an explication of 'e is evidence for h' should treat these cases differently. Intuitively, the same considerations that render e powerful evidence against h, when $0 < \mathrm{pr}_{\mathrm{old}}(h) < 1$, should have made e powerful evidence against h even if we had already been taking $\mathrm{pr}_{\mathrm{old}}(h) = 1$. BC's limited range of application thus becomes an argument against its appearing in an analysis of 'e is evidence for h'.

Of course, we might try to address this problem by expanding BC's scope. We might begin by making $\mathrm{pr}_{\mathrm{old}}(h|e)$ well defined even when $\mathrm{pr}_{\mathrm{old}}(e) = 0$, as Bruno de Finetti[11] does by holding $\mathrm{pr}_{\mathrm{old}}(h|e)$ to be unconstrained by unconditional probabilities when $\mathrm{pr}_{\mathrm{old}}(e) = 0$; $\mathrm{pr}_{\mathrm{old}}(h|e)$ is then no longer equal to $\mathrm{pr}_{\mathrm{old}}(h \,\&\, e)/\mathrm{pr}_{\mathrm{old}}(e)$ or to $\mathrm{pr}_{\mathrm{old}}(h)\mathrm{pr}_{\mathrm{old}}(e|h)/\mathrm{pr}_{\mathrm{old}}(e)$. We could then hold that even when $\mathrm{pr}_{\mathrm{old}}(e) = 0$, e's evidential relevance to h is given by BC–by comparing $\mathrm{pr}_{\mathrm{old}}(h|e)$ to $\mathrm{pr}_{\mathrm{old}}(h)$. But this move alone cannot resolve the problem. To explain why e is strong or weak evidence for or against h, we must still give some account of why $\mathrm{pr}_{\mathrm{old}}(h|e)$ takes on the value it does. Since this account fails to link $\mathrm{pr}_{\mathrm{old}}(h|e)$ to $\mathrm{pr}_{\mathrm{old}}(h)$, $\mathrm{pr}_{\mathrm{old}}(e)$, and the like in the usual way, the explanation of e's evidential relevance to h will be quite different from BC's explanation in cases where $0 < \mathrm{pr}_{\mathrm{old}}(h) < 1$ and $\mathrm{pr}_{\mathrm{old}}(e) > 0$. For example, BC is often cited as explaining what is right about hypothetico-deductivism–why the fact that h logically entails e can sometimes make e evidence for h: if h logically entails e, then (when $\mathrm{pr}_{\mathrm{old}}(h) > 0$) $\mathrm{pr}_{\mathrm{old}}(e|h) = 1$, and so $\mathrm{pr}_{\mathrm{new}}(h) =$

[10] For example, "contraction strategies" are discussed by Levi, pp. 59-62.

[11] *Probability, Induction, and Statistics* (New York: Wiley, 1972), pp. 81-82. See also, for instance, Levi, p. 219; and Earman, p. 36.

$\mathrm{pr}_{\mathrm{old}}(h|e) = \mathrm{pr}_{\mathrm{old}}(h)/\mathrm{pr}_{\mathrm{old}}(e)$, which (when $\mathrm{pr}_{\mathrm{old}}(e) < 1$) exceeds $\mathrm{pr}_{\mathrm{old}}(h)$. But this argument presumes $\mathrm{pr}_{\mathrm{old}}(e) > 0$; if $\mathrm{pr}_{\mathrm{old}}(e) = 0$, we would need an utterly different account of why $\mathrm{pr}_{\mathrm{old}}(h|e) > \mathrm{pr}_{\mathrm{old}}(h)$.

The fourth alleged problem for BC that I shall raise–and, arguably, the most fundamental–I shall call the *problem of incorrigibility*. If a scientist obeys BC after discovering e, and then again after discovering e', her $\mathrm{pr}_{\mathrm{new}}(h)$ is her $\mathrm{pr}_{\mathrm{old}}(h|e \,\&\, e')$. Therefore, a scientist who with each new observation alters her opinions by BC finds herself with opinions that were encoded all the while as conditional probabilities in her original probability distribution. There is a sense in which the probabilities in her most prior probability distribution are incorrigible; the scientist is forever trapped in this distribution, moving from point to point within it but never revising it. She is enslaved to commitments she undertook at the inception of her doxastic career, imprisoned by priors adopted (as Bas van Fraassen[12] says) "at her mother's knee." That these commitments are immune from revision is contrary to contemporary epistemological views and to scientific practice. Creative insights in science often occur precisely when a scientist, without receiving any new evidence or recognizing any logical relationship that had formerly escaped her, breaks out of what had been her most prior probability distribution–say, by deciding to take seriously a hypothesis to which she had been giving an exceedingly small (or zero) most prior probability, or to regard a hypothesis as capable of being confirmed by a class of evidence that she had formerly considered irrelevant to it.[13]

For instance, imagine a physicist of 1905 who is re-reading Albert Einstein's new paper "On a Heuristic Viewpoint concerning the Production and Transformation of Light."[14] Einstein examines the fact that Max Planck's equation governing the black-body spectrum can be derived from the hypothesis that light comes in discrete quantities (quanta) rather than continuous waves, and proposes that this equation's empirical correctness be regarded as evidence for any other prediction that can also be derived from the light-quantum hypothe-

[12] *Laws and Symmetry* (New York: Oxford, 1989) and "Rationality Does Not Require Conditionalization," in E. Ullmann-Margalit, ed., *The Israel Colloquium Series in the History, Philosophy, and Sociology of Science,* Volume V (Dordrecht: Kluwer, forthcoming).

[13] Levi lodges a similar objection against BC, recognizing that rationality leaves room for violations of "confirmational tenacity," as Levi aptly terms the perpetual grip of our most prior probability distribution (*op. cit.*, p. 82). (For more on this, see section II.) See also Howson and Alan Franklin, "Bayesian Conditionalization and Probability Kinematics," *British Journal for the Philosophy of Science,* XLV (1994): 451-66, here p. 459; Charles Chihara, "The Howson-Urbach Proofs of Bayesian Principles," in Eells and B. Skyrms, eds., *Probability and Conditionals* (New York: Cambridge, 1994), pp. 161-78, here pp. 171-73; and especially van Fraassen (*op. cit.*).

[14] *Annalen der Physik* (ser. 4), XVII (1905): 132–48.

sis. Suppose that on his first reading of this paper, our physicist shared the reaction of nearly all actual physicists in 1905: if the black-body spectrum behaves as if light is quantized, this is sheer coincidence–"nothing more than a curious property of light, without any physical significance."[15] It is not really a live option that light is quantized, considering the overwhelming success of the wave theory of light in accounting for a wide range of optical and electromagnetic phenomena. The light-quantum is just a formal device for deriving the correct black-body equation (as Planck[16] later reported himself to have held in 1905). Accordingly, it would be utterly coincidental if any other phenomenon were as if light is quantized. So the black-body equation's success does not confirm the other equations (governing the photoelectric effect, the photoionization of gases, and photoluminescence) that Einstein uses the light-quantum hypothesis to derive. This, I said, was our physicist's initial reaction. But suppose that on re-reading Einstein's paper, our physicist is more impressed by certain considerations which he noted before, but which (he now believes) he originally failed to accord their proper weight. For example, he is struck more forcibly by the paper's opening thought: that none of the remarkable successes of the classical theory of light (reflection, refraction, diffraction, interference, and so on) concerns light absorption or emission, which are central to each of the phenomena with which Einstein is concerned (the black-body spectrum, photoelectricity, and so on). Although he knew this point before re-reading the paper, our physicist decides that he failed then to give it due regard. He therefore re-evaluates certain of his opinions–for example, raising his pr(light-quantum hypothesis) and setting pr(Einstein's photoelectric-effect equation | Planck's black-body equation) > pr(Einstein's photoelectric-effect equation), whereas he had formerly set these equal. He is not taking into account evidence (or even logical truths) that he did not know or take into account before, nor is he inventing some new hypothesis. Rather, he regards his new probabilities as the *right* ones for him to assign–the ones he *should* have assigned earlier, considering what he knew then. (I shall elaborate this idea of rightness in the next section.)

Now perhaps it is not initially evident how the problem of incorrigibility is a problem for BC. Our physicist has shifted his opinion, but did not discover any new evidence regarding light quantization, whereas BC purports to cover only shifts involving new evidence. Yet

[15] Abraham Pais, *Subtle Is the Lord* (New York: Oxford, 1982), p. 337.

[16] From a 1931 letter to R.W. Wood, in A. Hermann, ed., *Fruhgeschichte der Quantentheorie*, Claude W. Nash, trans. (Cambridge: MIT, 1971), pp. 23-24.

we might just as well have imagined an example where a scientist's creative insight–her breaking out of her most prior probability distribution–occurred together with her discovery of some new evidence, and so her shift of opinion falls within BC's jurisdiction, but violates BC. Indeed, our physicist's shift of opinion with regard to light quantization surely happened to be accompanied by his learning *something*–perhaps that the dog is barking–and so he updated his opinions in violation of BC. Thus, I believe that it is permissible for updates to violate BC.[17]

We might amend BC by explicitly restricting its application to cases where we should not break out from our most prior probability distribution. But since $pr_{new}(h) = pr_{old}(h|e)$ follows from our retaining our most prior probability distribution (in particular, from $pr_{new}(h|e) = pr_{old}(h|e)$), BC is trivialized by being so amended; it then says merely that we should update by $pr_{new}(h) = pr_{old}(h|e)$ in cases where we should. Instead, we might amend BC by restricting its application to cases where we have no reason to depart from our most prior probability distribution, and by offering further rules that govern what count as good reasons for so departing.[18] But BC can then account only for those shifts of opinion where the agent should remain within her most prior probability distribution. This apparently threatens all of the familiar Bayesian accounts of various features of confirmation, since these accounts will not apply unless the agent must remain within her most prior probability distribution, and until we have developed some further rules, we cannot explain when this is. I shall argue, however, that although our shifts of opinion need not leave us within our most prior probability distribution, the steps in an argument justifying our current opinion should do so. BC can then be used straightforwardly–without amendment–to explicate '*e* is evidence for *h*', thereby preserving the familiar Bayesian accounts intact.

II

Here, then, we have four problems that arise when we try to use BC to explicate '*e* is evidence for *h*'. I shall suggest that these problems fail to arise if our explication depends not on BC's governing how we should update our opinions from one moment to the next, in light of new evidence, but instead on BC's governing how we should justify our current opinions by appealing to evidence that we have already accumulated. In showing an audience how her current degree of confidence in *h* constitutes a justified response to the evidence

[17] Again, here I concur with van Fraassen, *op. cit.*

[18] Levi (*op. cit.*) allows confirmational commitments to be revised in various ways depending upon the context of inquiry. He offers an account of how we ought to revise our "confirmational commitments" in certain important contexts.

that she already has on hand, a scientist explains the bearing of various pieces of evidence on *h*, but does not change her opinions in the course of doing so.

She begins her justificatory argument, I suggest, by appealing to a probability distribution that other scientists should recognize as unbiased in the relevant respect. She is entitled to begin with this distribution as a "free move"–that is, without justifying it.[19] She then justifies her current degree of confidence in *h* by invoking various pieces of evidence that have already been discovered, and by using them successively to "update" the initial, unbiased opinion until all of the relevant evidence has been considered, at which point her current degree of confidence in *h* should be reached. Once again, I put scare-quotes around 'update' because the scientist is not changing her degrees of belief in the course of this exercise. Rather, she is justifying the opinion that she was already holding when she began to demonstrate that she was entitled to hold it. (In the next section, I shall argue that this "updating" must accord with BC.)

For instance, consider how a weather forecaster might justify her current 80% confidence in the hypothesis that it will rain and not snow tomorrow. She begins her justificatory argument with an "unbiased" probability distribution, which (let us suppose) assigns 20% confidence to 'It will rain and not snow tomorrow' (hypothesis h_1), 10% confidence to 'It will rain and snow tomorrow' (h_2), 5% confidence to 'It will snow and not rain tomorrow' (h_3), and so on. In the first step of her argument, she submits her first piece of evidence: that today's barometric pressure was 29 torr. This evidence, she explains, warrants "updating" her opinions in a certain way; among her "intermediate conclusions" at this first step are $\mathrm{pr}(h_1) = .6$, $\mathrm{pr}(h_2) = .10$, and $\mathrm{pr}(h_3) = .01$. Today's barometric pressure, then, is evidence for h_1 and against h_3, while it counts neither for nor against h_2. In the second step of her justificatory argument, she takes into ac-

[19] To invoke this distribution's "impartiality" is obviously to issue a giant promissory note. Of course, its impartiality cannot depend on its somehow according all conceivable competing hypotheses equal weight. Later, I say a bit more about impartiality, but to pursue it much further would amount to discussing the problem of the priors, which I shall not do. Let me emphasize that, if my position ultimately requires an account of what makes some probability distribution able to serve as the initial distribution, so does the view that BC governs the revision of opinion in the face of new evidence. I criticized that view as unable, for various reasons, to explicate '*e* is evidence for *h*', but not for presupposing an answer to the problem of the priors; the criticisms I reviewed would still apply even if we had some account of why certain opinions constitute a reasonable place to begin one's doxastic life. My purpose here is to find some way to preserve BC's apparent successes while avoiding the problems I have mentioned, not to resolve the problem of the priors.

count another piece of the evidence–perhaps today's temperature–she has on hand. She explains the bearing of this evidence on the leading hypotheses; among her intermediate conclusions at this second step are $\mathrm{pr}(h_1) = .7$, $\mathrm{pr}(h_2) = .002$, and $\mathrm{pr}(h_3) = .001$. Today's temperature is evidence for h_1 and against h_2 and h_3. She continues in this fashion until she has submitted all of the evidence she has acquired and deems relevant. For the justificatory argument to succeed, she must at this final step have reached $\mathrm{pr}(h_1) = .8$, her current opinion.

Changing your opinion upon ascertaining new evidence is utterly distinct from justifying your current opinion by appealing to evidence that you have already accumulated. The justification of a scientist's current opinion is not necessarily the history of her arrival at that opinion, even if she arrived at that opinion justly. For instance, the scientist may at one time have revised her opinion regarding the hypothesis by considering new evidence that she later decided was false. In the argument that now justifies her degree of confidence in that hypothesis, there is no step in which her opinion is "updated" in light of this evidence, since she now believes this evidence mistaken. Furthermore, her current justification may differ from her earlier justification even if the opinion being justified and the evidence justifying it are unchanged. She may have changed her mind in the meantime regarding whether certain observations count as evidence for or against the hypothesis, or how powerfully they count. Her current justificatory argument takes each successive piece of evidence into account in precisely the manner that she now believes proper.

The distinction between updating your opinions upon receiving new evidence and showing how the evidence on hand justifies your current opinions should not be confused with various other, more familiar distinctions, such as that between the "context of discovery" and the "context of justification." Admittedly, both distinctions have a diachronic side concerned with arriving at something and a synchronic side concerned with justifying something. A crucial aspect of the traditional distinction between the two contexts, however, is that when we discuss some hypothesis in the context of discovery, we make no normative judgments; we describe how (for example) a conversation, a piece of evidence, a dream, or a heuristic procedure led to the formulation of the hypothesis. In contrast, *both* sides of my distinction are normative: how you *ought* to revise your past opinions in response to new evidence versus how your current evidence *warrants* your current opinions. The context of discovery includes various influences, such as dreams, that appear on neither side of my distinction, since they play no justificatory role.[20]

One way to grasp the role in justificatory arguments that (I shall ultimately suggest) BC plays is through the distinctions that Levi carefully draws among the various roles that BC might play in connection with updating opinions. Levi presumes that all rational agents adopt "confirmation commitments" that specify, for each deductively closed set of claims, the opinions that the agent is committed to holding in the event that the claims to whose truth she is committed are exactly the members of that set (*op. cit.*, p. 79). Levi argues that our confirmational commitments at a given moment are constrained by "confirmational conditionalization"–roughly, that we are committed to updating our opinions in accordance with BC (*op. cit.*, pp. 81, 222). Again very roughly: for a consistent expansion of our current full beliefs, formed by adding e to those beliefs and then rendering the result deductively closed, "confirmational conditionalization" requires that we be committed to our $\mathrm{pr}_{\mathrm{old}}(h|e)$ being equal to our $\mathrm{pr}_{\mathrm{new}}(h)$. If our confirmational commitments are constrained by confirmational conditionalization and remain unchanged upon a consistent expansion of our current full beliefs, then our new opinions are dictated by BC. Such a shift of opinion is what Levi calls a "temporal credal conditionalization." But Levi denies "confirmational tenacity" (thereby recognizing what I have called the problem of incorrigibility) (*op. cit.*, p. 82); that is, he denies that a rational agent must always remain faithful to her former confirmational commitments, and so carry out their dictates on the occasion of (say) a consistent expansion of her beliefs in light of an observation. Rather, she may have good reason for revising her confirmational commitments.[21] In that case, her shift of opinion, upon a consistent expansion of her current beliefs, will not generally be a temporal credal conditionalization.[22] Now, rather than consider an agent's shift of opinion from one moment to the next, let us consider her transitions from one step to the next in a justificatory argument. At each step, she has arrived at certain intermediate conclusions along with (let us suppose) certain "confirmational

[20] Some authors have argued that normative judgments can be made regarding a scientist's means of thinking up hypotheses to entertain. But even if there is a "logic of discovery," BC could play no role in it, since BC operates only on hypotheses that have already been formulated.

[21] See footnote 18.

[22] Levi notes that, even if the agent retains her confirmational commitments, it does not follow that all of her shifts of opinion are temporal credal conditionalizations, since perhaps a legitimate revision may be occasioned by a contraction rather than an expansion of her beliefs, or by new evidence that does not involve learning the truth of any claim. See footnote 10.

commitments" that specify, for any claim e that is consistent with all of the full beliefs already arrived at by that step, the opinions that represent the right way to take e into account, should e be the further evidence submitted in the next step of the justificatory argument. The analog of "confirmational conditionalization" requires that at each step, the agent's confirmational commitments be governed by BC, in the following sense: if the agent is required to honor her confirmational commitments in moving to the next step in the justificatory argument (the analog of "confirmational tenacity"), and if every transition from one step to the next results from taking account of a further piece of evidence e that is consistent with all of the full beliefs arrived at thus far in the argument, then the analog of "confirmational conditionalization" requires that the agent's intermediate conclusion $\mathrm{pr}_{\mathrm{new}}(h)$ at the next step equal her intermediate conclusion $\mathrm{pr}_{\mathrm{old}}(h|e)$ at the current step. This "update" is the analog of a temporal credal conditionalization. In effect, I defend these analogs of confirmational conditionalization and tenacity, and argue that all transitions from one step to the next in a justificatory argument are analogs of temporal credal conditionalizations.

In other words, I have suggested (in connection with the problem of incorrigibility) that a scientist is not obliged to revise her opinions, *from one stage in her work to the next*, in accordance with BC. Nevertheless, I shall argue that *at each stage in her work*, BC must govern the steps in the arguments by which her opinions then are justified. In that event, BC would be in a position to account for various features of confirmation without encountering the four problems that I have mentioned.

For instance, the problem of old evidence–as a difficulty for using BC to explicate 'e is evidence for h'–is avoided if this explication applies BC not to updatings, but to the transitions between the steps in justificatory arguments. Although our confidence in h was never *updated* in light of e's discovery, since e was discovered before h was formulated, the argument by which we *now justify* our $\mathrm{pr}(h)$ appeals to e. That argument begins with a probability distribution in which h is assigned some nonzero probability and e is assigned a probability less than 1 (since this distribution did not take account of any of the evidence to be submitted in the course of the argument). So in the context of a justificatory argument, none of the evidence being presented for (or against) h is "old."

Notice that this response to the problem of old evidence is similar to but distinct from the appeal to counterfactual degrees of belief (which I mentioned earlier). A prior probability in a justificatory argument is like a counterfactual degree of belief in that it may be a fictitious state of opinion; perhaps at no moment did our opinions match this prior probability distribution. As in the appeal to counterfactual degrees of belief, the strategy is to argue that there is a reasonable prior that does not construe the old evidence as old. But unlike the counterfactual approach, the prior probability distribution in a justificatory argument is not beholden to the historical facts that determine the confidence that we would have placed in the hypothesis, had it been formulated sometime earlier. These historical facts are irrelevant to our current judgments of whether *e* is evidence for *h*. A given prior probability distribution is relevant to whether *e* is evidence for *h* not because, in the actual world or some possible world, we hold this probability distribution and then raise our confidence in *h* when updating our opinions upon discovering *e*. Rather, it is relevant because this probability distribution figures in the arguments by which we justify our opinion regarding *h*.

It might be objected that I have not removed the problem of old evidence, but merely presupposed that there is a solution to it–that there is a reasonable prior that does not construe the old evidence as already known. But this charge is unfair. Once we shift our attention from updates to justificatory arguments, the temporal element is removed, and so old evidence presents no peculiar problem. The *oldness* of the old evidence presents no special reason for it to be depicted by the initial probability distribution in a justificatory argument as having already been ascertained; a piece of evidence submitted in the course of the justificatory argument, whether or not it was discovered before the given hypothesis was formulated, was perforce found before the justificatory argument was offered. So there is no longer any problem of old evidence. Of course, there remains the question: Whence comes the initial probability distribution in the justificatory argument? But we have not merely exchanged the problem of old evidence for this problem, since this problem was already with us as a version of the problem of the priors.[23]

The second problem I raised–how to use BC to explicate *e*'s serving as evidence for (or against) *h*, if *h* is already believed (or rejected) by the time *e* is discovered–is again not a problem once we stop trying to find an updating to which BC applies, and instead apply BC to the step where *e* is taken into account in a justificatory argument. After we adopt (or reject) *h*, we continue to be committed to arguments that justify our doing so. Suppose that even before discovering *e*, we justly adopted *h*; we already had sufficient evidence to fund an argument

[23] See footnote 19.

justifying belief in *h*. Then *e*'s discovery may nevertheless enable us to give a new argument for believing *h*, one step of which involves taking *e* into account. If that step involves an increase in our confidence in *h*, then it is correct to say (in connection with this argument) that we take *e* as evidence for *h*. In short, there is only one history of the shifts in our opinions upon discovering new evidence, but we may have multiple arguments for our current degree of confidence in *h*. So even if *e*'s discovery resulted in no further shift in our confidence in *h*, nonetheless *e* may count as evidence for (or against) *h*, so long as this is cashed out in terms of *e*'s role in our justificatory arguments.

The third problem I raised–that BC cannot explain what makes *e* evidence against *h*, if *h* is initially believed and then *e* leads us to reconsider–likewise can be avoided if BC is applied not to *e*'s updating our confidence in *h*, but to *e*'s role in an argument justifying our current degree of confidence in *h*. That is because no claim accepted at one step in a justificatory argument is rejected at a subsequent step. For instance, if a scientist initially accepts a given claim as an observation report but later rejects it, she thenceforth ceases to submit that claim as an observation report anywhere in her justificatory arguments. She does not cite it at one stage in an argument, as evidence relevant to various hypotheses under consideration, and then take it back at a subsequent stage. As Thomas Kuhn[24] emphasizes, the justificatory argument in a scientific paper is "cleaned up"; it does not double-back on itself in the way that a scientist's actual doxastic history might.

Similarly, the problem of incorrigibility is avoided when BC is applied to the steps in a justificatory argument. Our 1905 physicist changed her mind about whether the correctness of Planck's blackbody equation counts as evidence for Einstein's photoelectric-effect equation. In contrast, once having submitted *e* at some step in a justificatory argument as evidence for (or against) *h*, we cannot at some other step in the argument revise our opinion of whether *e* counts as evidence for or against *h*. The argument, I shall suggest, operates within a fixed most prior probability distribution (though we may, of course, later reject the argument). For example, suppose you are justifying your current 75% confidence in *h*. After the first few steps in your argument, you have taken into account a certain portion of your current evidence and reached $pr(h) = .4$ as your confidence so far. After the next few steps, you have taken into account more of your current evidence and reached $pr(h) = .8$ as your intermediate conclusion. At this step, you cannot reject your "earlier" degree of confidence as having been the wrong intermediate conclusion to

[24] *The Essential Tension* (Chicago: University Press, 1977), p. 327.

have reached from the evidence under consideration at that "earlier" step. *Since the intermediate conclusions at "earlier" steps in a justificatory argument are the basis for the intermediate conclusions at "later" steps, a "later" step cannot involve the rejection of an "earlier" intermediate conclusion as having been an inappropriate conclusion to have reached from the evidence under consideration at that step.* The opinions appearing as intermediate conclusions at a given step of the justificatory argument must be regarded *throughout the argument* as having been merited by the evidence taken into account at, or prior to, that step.

I shall now try to spell out this view by using an argument from calibration to show that the steps in a justificatory argument must accord with BC.

III

A standard way of arguing for BC is by a Dutch Book argument. A *Dutch Book* is a series of bets, each bet deemed fair by the bettor at the time she makes it, such that the bettor will lose money no matter what the outcomes of the events that are the subjects of these bets. In a *synchronic* Dutch Book, the bettor makes all of the bets at once; in a *diachronic* Dutch Book, she makes them at different times. The idea is that, if the bettor's degrees of belief regarding the outcomes of various events are such that the bookmaker can design bets on these outcomes which the bettor will take (that is, that she will deem fair) but which must necessarily result in the bettor's suffering a net loss, then the bettor must have made a logical error in the management of her opinions. It is required that the bookmaker be able to design the Dutch Book even though the bettor and bookmaker know exactly the same facts. For instance, if the bettor uses a rule for updating her opinions upon receipt of evidence, then since she must know the rule she uses, the bookmaker must know that rule, too, and if the bettor does not know the outcomes of the events on which she has bet, then the bookmaker must not know them either. (Obviously, if the bookmaker knows the outcome of the event on which the bettor is betting, and the bettor does not, then it signifies no logical flaw in the bettor's management of her opinions that the bookmaker can construct a bet which the bettor believes fair but from which she cannot profit.) It was shown (by John G. Kemeny, R. Sherman Lehman, and Abner Shimony, with key contributions from Frank Ramsey and de Finetti) that you are vulnerable to a synchronic Dutch Book if and only if your degrees of belief at a given time violate the axioms of the probability calculus, and it was shown (by David Lewis) that you are vulnerable to a diachronic Dutch Book if and only if you update your degrees of belief by using a rule other than BC.

But we cannot use a diachronic Dutch Book argument to show that the steps of a justificatory argument must conform to BC. We might worry that a Dutch Book scenario is not an apt metaphor for an agent's offering a justificatory argument, since an agent does not receive new evidence and revise her opinions in the course of justifying her current opinions. But even setting this worry aside by treating the "updates" in a justificatory argument as like diachronic shifts of opinion, there is a further problem: a Dutch Book argument cannot show that we must update according to BC, since we can always frustrate the bookmaker by using no rule at all to update our opinions. The Dutch Book scenario presumes that, if we do not know in advance how we would revise our opinions under various circumstances, then the bookmaker does not know this either, and so cannot construct a Dutch Book against us. Using no rule is not an empty threat; we saw our 1905 physicist use no rule in changing his mind about whether the success of Planck's black-body equation confirms other equations derivable from the light-quantum hypothesis. Accordingly, van Fraassen[25] maintains that, although a Dutch Book argument demonstrates that the only rational *rule* for updating is BC, it permits us rationally to depart from BC so long as we employ no rule at all for updating our opinions. Why, then, is a justificatory argument unsuccessful if its steps depart from BC?

Let us begin with a logically prior question. Suppose we present a justification of our current degree of confidence in a hypothesis, proceeding from an "unbiased" most prior distribution by a sequence of updates. Each update responds to one piece of evidence that we have accumulated, and the sequence of updates continues until all of the evidence on hand that we deem relevant has been considered. Why does this exercise possess the power to justify our current opinion? Why do we succeed in justifying our current opinion by showing that it is guided by the degrees of confidence in the prior step, which are guided by the degrees of confidence in the yet prior step, and so on back to the "unbiased" most prior probabilities? Why does guidance by those intermediate conclusions and initial opinions matter?

It matters, I suggest, because we believe the intermediate conclusions and initial opinions to be "perfectly calibrated." Consider a weather forecaster who has 80% confidence that it will rain tomorrow. Suppose that she bases this opinion on the fact that it rained today, the high temperature today was 60 degrees, and so on. Any other day where it rains on the previous day, the high temperature on the previous day is 60 degrees, and so on–in other words, any other day that is

[25] *Laws and Symmetry*, p. 174; "Rationality Does Not Require Conditionalization."

similar to tomorrow, in all of the respects that grounded the forecaster's opinion regarding rain tomorrow–I shall term *relevantly similar* to tomorrow. Suppose that the weather forecaster knows regarding some other day only that it rained on the previous day, the high temperature on the previous day was 60 degrees, and so on. Then the weather forecaster should have 80% confidence that it will rain on that day, since she has exactly the same reasons in the two cases for believing that it will rain. In assigning 80% confidence to the claim that it will rain tomorrow, the forecaster is perfectly calibrated if and only if it rains on exactly 80% of the days that are relevantly like tomorrow. If it rains on more than 80% of those days, then the forecaster's subjective degree of belief is too low; if it rains on less than 80% of those days, then her opinion is too high. If she is perfectly calibrated, then her opinion is right; it is just what it ought to be, considering her reasons for it (that is, her sense of relevant similarity).[26]

In a justificatory argument, we base our intermediate conclusion '$\mathrm{pr}(h) = r$' on certain considerations that we have already submitted by that step. Suppose that we consider some other hypothesis k, discussed at that step, to be relevantly similar to h, in that we take the considerations on which we base our intermediate conclusion '$\mathrm{pr}(k) = s$' at that step to be relevantly similar to those on which we base '$\mathrm{pr}(h) = r$' at that step. That is, we take ourselves to have exactly the same kinds of reasons for our degree of confidence in k as we do for our degree of confidence in h. (Accordingly, $r = s$.) The intermediate conclusions at a given step are "perfectly calibrated" exactly when for each of these conclusions '$\mathrm{pr}(h) = r$', r is the fraction of true hypotheses among the hypotheses that at this step we deem relevantly similar to h. I suggest (as a first approximation) that we believe the intermediate conclusions at a given step of a justificatory argument to be perfectly calibrated, since this would explain *why* we believe that we ought to be guided by those conclusions when we move to the next step, taking some further evidence into account. (In a moment, I shall examine just *how* we should be so guided.)

Let me elaborate. Suppose we have no confidence in the calibration of a weather forecaster's predictions. For example, she has 80% confidence that it will rain tomorrow, but our opinions regarding the frac-

[26] For a similar conception of an opinion's rightness, see van Fraassen, "Calibration: A Frequency Justification for Personal Probability," in R.S. Cohen and L. Laudan, eds., *Physics, Philosophy, and Psychoanalysis* (Boston: Reidel, 1983), pp. 295-319, here p. 301; *Laws and Symmetry*, p. 177. I shall qualify this idea shortly. As van Fraassen emphasizes, this conception of an opinion's rightness does not suggest that our subjective degree of confidence in rain tomorrow is actually to be understood as a claim concerning the objective chance of rain then.

tion of rainy days among the days that she takes to be relevantly like tomorrow are no different from our opinions regarding the fraction of rainy days among the days that she takes to be relevantly like the day after tomorrow, even though she has 80% confidence that it will *not* rain on that day. Then we do not take her forecasts seriously; we give them no weight in guiding our opinion regarding whether it will rain.[27] In a justificatory argument for our current opinion, by contrast, each step's intermediate conclusions are guided by the previous step's; we take those intermediate conclusions seriously when we take some further evidence into account. If we change our minds about the right intermediate conclusions to draw at a given step, then we change that step, adjusting the conclusions there accordingly. We do not retain the step with the intermediate conclusions that (we now judge) fail to accord certain considerations their proper weight, and add to the argument a further step to negate the effect of these conclusions on subsequent steps. To include a step where no additional evidence is submitted, but new intermediate conclusions are drawn to replace the old, would be in effect to start the argument over in the middle by setting aside the intermediate conclusions drawn in the earlier step as having been mistaken there. (I present an example of this in a moment.) The simplest rational reconstruction of why the intermediate conclusions in a justificatory argument are taken seriously throughout the argument, each step's guiding the next step's, is that the agent believes each step's intermediate conclusions to be perfectly calibrated.[28]

In contrast, we *can* rationally change our opinions *from one moment to the next* without ascertaining any new evidence. When we decide that our earlier opinions failed to accord certain considerations their proper weight, we decide that those opinions were not the right ones for us to have held, considering the evidence we then knew. We are taking into account no evidence (or even logical truth) that we formerly failed to take into account; rather, we are changing our minds about *what it takes* for us to take proper account of some consideration that we already knew.

So the key difference between updatings from one moment to the next and "updatings" from one step in a justificatory argument to the

[27] For a similar view, see Maurice Fréchet, "Les Définitions Courantes de la Probabilité," *Revue Philosophique*, CXXXVI (1946): 129-69, here p. 144.

[28] Here is a less simple but perhaps more realistic alternative. Rather than require that the agent believe r to be the fraction of true hypotheses among the hypotheses relevantly like h, require that r be the agent's *expectation value* for that fraction. That is, require that $r = \Sigma \mathrm{pr}(r_i) r_i$, where $\mathrm{pr}(r_i)$ is the agent's subjective probability that r_i is that fraction. This modification is compatible with the subsequent argument, *mutatis mutandis*.

next is that in the justificatory case, but not in the diachronic case, we must use the earlier opinions to guide their revision; we deem the later opinions to be right on the basis of deeming the earlier opinions to have been right. If we decide that some intermediate conclusion was not right, then we change that step, giving ourselves a new justificatory argument. But if we change our minds about whether our opinions at an earlier moment accorded proper weight to the evidence then known, we cannot go back and change which opinions we held in the past.

To see how the later steps of a justificatory argument are supposed to be guided by the earlier steps, so that the earlier steps must be considered throughout the argument to have been right, consider this example. Suppose that a weather forecaster sets out to justify her 80% confidence in rain tomorrow. She begins her argument with an "unbiased" most prior probability distribution. She then considers a piece of evidence: that today's barometric pressure is 29 torr. She assigns tomorrow to a reference class: the set of days each of which is preceded by a day with barometric pressure close to 29 torr. She judges the cases belonging to the same reference class to be "relevantly similar," that is, similar in all of the respects on which she bases her intermediate conclusion at this step regarding whether it will rain tomorrow. Accordingly, in the intermediate conclusions at that step, she assigns the same degree of confidence to rain on each of these days–let us say 50%. She then considers another piece of evidence: whether or not it rained today. She assigns tomorrow to a narrower reference class (taking a different set of cases to be relevantly similar) and accordingly adopts a new intermediate conclusion. Now suppose that, if her second piece of evidence is that it rained today, the weather forecaster is prepared to make her second step's intermediate conclusion 'pr(rain tomorrow) = .2', and if her second piece of evidence is that it did not rain today, the weather forecaster is prepared to make her second step's intermediate conclusion *again* 'pr(rain tomorrow) = .2'. This would be tantamount to her changing her confidence in rain tomorrow from one step to the next without any additional evidence as motivation; plainly, she is not motivated by taking into account that it did rain, as the opposite discovery would have made no difference to her second step's intermediate conclusion, and she is not motivated by taking into account that it either rained yesterday or it did not, as this is a trivial truth and so does not count as new evidence. To go in this way from 50% to 20% confidence in rain tomorrow is to fail to take the first step's intermediate conclusion seriously; she is not being guided by her first step's intermediate conclusions in reaching her second step's. That is not the way a justificatory argument proceeds.

The same flaw occurs even if the two possible intermediate conclusions ('pr(rain tomorrow) = ...') in the second step are not identical, if they are both greater than .5 (the first step's intermediate conclusion) or both less than .5. In either case, the agent is not taking the first step's intermediate conclusion seriously. For instance, if both of the second step's possible intermediate conclusions assign less than 50% confidence to rain tomorrow, then she cannot regard it as being right to assign 50% confidence in the first step; she was prepared to lower her pr(rain tomorrow) in the second step no matter what the evidence concerning rain today, so she must regard the first step's 50% assignment as excessively high. This logical flaw is manifested in the fact that the agent cannot with consistency regard her first step's intermediate conclusions as perfectly calibrated and her second step's intermediate conclusions as perfectly calibrated. For if rain occurs in less than 50% of the days following 29-torr rainy days, and occurs in less than 50% of the days following 29-torr nonrainy days, then it cannot be that rain occurs in 50% of the days following 29-torr days.[29] In failing to take her intermediate conclusions seriously, the weather forecaster betrays her belief that they are not calibrated.

Let me elaborate the notion of a "reference class" as it pertains to calibration. For me to believe my intermediate conclusion 'pr(h) = r' to be perfectly calibrated is for me to believe that r is the fraction of truths among the hypotheses that in this step I believe "relevantly similar" to h (and so to which I assign the same degree of confidence as I do to h). Thus, my belief in the perfect calibration of my intermediate conclusion is meaningful only if there is a meaningful sense of hypotheses "relevantly like" h. Readers may now recall the notorious *problem of the reference class* (or *problem of the single case*) as it afflicts various frequentist analyses of probability: If 'the probability of this coin toss landing heads is 50%' just means that 50% of a large collection of tosses like this one would land heads, what count as tosses *like* this one? Despair over the reference-class problem has often been cited as an argument against any such analysis of probability. But I am not offering an *analysis* or *interpretation* of the probability assignments in justificatory arguments. The opinions that we believe perfectly calibrated are not opinions *about* the fraction of truths

[29] My flawed weather forecaster is intentionally similar to van Fraassen's Piero in "Belief and the Problem of Ulysses and the Sirens," *Philosophical Studies*, LXXVII (1995): 7-37, here p. 15. His example involves a change in opinion from one moment to the next rather than from one step to the next in a justificatory argument. What van Fraassen calls the forecaster's lack of "integrity" I elaborate as her failure to take seriously her first step's intermediate conclusion.

among relevantly similar hypotheses; they are subjective probabilities. Moreover, the notion of calibration that I am employing is not afflicted by any problem of the reference class. The hypotheses that I take in this step to be "relevantly similar" to h are exactly the hypotheses that are similar to h in all of the respects on which I base my intermediate conclusions at this step. In other words, they are the hypotheses that I take myself at this step as having the same sorts of reasons for believing as I do for believing h. For the hypothesis of rain tomorrow, it is easy to imagine what it might take at a given step for a hypothesis to be relevantly similar. For another sort of hypothesis, it may be more difficult to characterize the reasons we have for assigning to it the degree of confidence we do.[30] But to make my point, I do not need to offer an account of these reasons. I need presume only that we must have reasons for assigning to hypotheses the degrees of confidence we do (even if we have not explicitly identified these reasons). Unlike the frequentist interpretations of probability I mentioned, talk of perfect calibration does not presuppose some *objective* sense of two hypotheses being "relevantly similar." The characteristics by which *we are guided* in confirming the hypothesis determine its reference class. We take ourselves as being guided by some factors or other (which, once again, we may be able to express in only a rough and ready fashion). I return to this point in the next section.[31]

[30] Notice that I need not regard every hypothesis to which (at this step) I assign the same degree of confidence as I assign h to be "relevantly similar" to h. (van Fraassen makes the same point in "Calibration: A Frequency Justification," p. 303; see also the next footnote.) I may consider my reasons for assigning r to a hypothesis predicting rain to be quite different from my reasons for assigning r to a hypothesis predicting that a heart-transplant patient will survive for five years after his transplant.

[31] van Fraassen offers one way to cash out this notion of the reference classes in terms of which our calibration is assessed. In "Calibration: A Frequency Approach," he says (pp. 302-05) that the reference class for the claim '*x is A*' (perhaps '22 May 1998 will be rainy in Seattle') consists roughly of '*y is A*', '*z is A*', and so on, where for every property *B*, our pr(*x is B*) = pr(*y is B*) = pr(*z is B*), and so on. By restricting the reference class of '*x is A*' to claims attributing *A* to other entities, van Fraassen prevents the reference class for '*x is A*' from consisting of '*x is A*' along with arbitrarily many claims of the form 'The *n*th toss of the die lands ace'. This is important for the following reason:

> [S]uppose that I first state my probability for rain as 1/6 and then you ask me about one thousand tosses of a fair die for the probability of ace and I say 1/6 each time. On the total set of 1001 questions, my personal probability will probably be quite well calibrated, but that reveals nothing about the reasonableness of my initial judgment about rain. To see the problem in acute form, let this first judgment be replaced by two: adding to it also the judgment that the probability of there being no rain equals 1/6 as well. Calibration on the total of 1002 propositions will be quite good, whereas there is something drastically wrong with my probabilities for the first two.
>
> So the possibility of ever better calibration which we require, must be on extensions of the initial set of propositions which are in a relevant sense *like* the original ones (pp. 304-05).

van Fraassen's thought seems to be that a rational agent must have the same reasons for her opinion regarding *j* as for her opinion regarding '*x is A*' if *j* attributes

I have suggested that an argument deriving our current degree of confidence in a hypothesis from an "unbiased" most prior distribution, by a sequence of updates that eventually takes all of our evidence into account, possesses the power to justify our current opinion because we are entitled to believe that this argument's intermediate conclusions are perfectly calibrated. We are entitled–without having to submit any argument–to believe the most prior distribution to be perfectly calibrated; that this is a free move is part of what it is for the most prior distribution

---

A to some other entity y and if she has the same confidence in '*y is A*' as in '*x is A*', the same confidence in '*y is B*' as in '*x is B*', and so on. van Fraassen refuses to cash out the reference class by appealing directly to our having *the same reasons* for our confidence in j as for our confidence in '*x is A*'; he is willing to appeal only to features of our probability distribution.

But this self-imposed limitation leads van Fraassen to resort to the heavy-handed measure of requiring the reference class for '*x is A*' to consist exclusively of claims attributing A to other entities. It is not at all clear to me how this analysis applies when the claim in question is not of the monadic form '*x is A*'–for example, when the claim is quantified. How do we assess whether a nineteenth-century physicist's opinion of Newton's gravitational-force law is calibrated?

I am not as squeamish as van Fraassen is about carving out h's reference class by appealing directly to our reasons for holding our opinion regarding h. I have no way to read these reasons off of our probability distribution. Analyzing what it means to say that our opinion regarding h is *based* on a given consideration is a classic epistemological problem. See, for instance, Keith Lehrer, *Knowledge* (New York: Oxford, 1974), and George Pappas, "Basing Relations," in Pappas, ed., *Justification and Knowledge* (Boston: Reidel, 1979), pp. 51-63. I assume nothing about what our reasons are, only that we must have reasons. (See section IV.) They presumably relate to our probability distribution in something like the following way. Our reasons for assigning a certain degree of confidence r to h consist of our belief that h possesses various properties $\alpha, \beta, \ldots$, where these properties could logically possibly hold of any number of (logically inequivalent) claims. Let Γ be a dummy variable standing for a claim. Then consider our $\mathrm{pr}(\Gamma \mid \Gamma$ possesses $\alpha, \beta, \ldots)$. In other words, for an unidentified claim, this is our degree of confidence in it, given various facts about it. So this might be $\mathrm{pr}(\Gamma \mid \Gamma$ says that it will rain on a certain day, this day follows a day where the high temperature was 80 degrees and the barometric pressure was 30 torr,...). Our belief that h possesses $\alpha, \beta, \ldots$ can be our reason for assigning r to h only if our $\mathrm{pr}(\Gamma \mid \Gamma$ possesses $\alpha, \beta, \ldots) = r$, and for any further claim g that we believe, our $\mathrm{pr}(\Gamma \mid g, \Gamma$ possesses $\alpha, \beta, \ldots) = r$. (For some general remarks in a similar spirit, see Wesley Salmon's discussion of the kinds of properties the possession of which by a hypothesis can serve as a basis for a plausibility judgment–in *The Foundations of Scientific Inference* (Pittsburgh: University Press, 1967), pp. 124-26.)

The above constraint on h's reference class ensures, for instance, that for any k in that class, $\mathrm{pr}(k) = \mathrm{pr}(h)$. This is also a feature of van Fraassen's account. But this constraint does not preclude the reference class of '22 May 1998 will be rainy in Seattle' from consisting of arbitrarily many claims of the form 'The nth toss of the die lands ace'. I have presented no *logical* bar to someone's having the following reason for assigning a degree of confidence of 1/6 to j: either j concerns rain in Seattle on 22 May 1998 and the high temperature on the previous day was 80 degrees,..., *or* j concerns a fair die landing ace. *We* presumably have different reasons for our opinion regarding the rain claim, however, than for our confidence in the die claim. (To say just what makes this so, we would need to identify the "basing relation.")

to be "unbiased."[32] Each step of the justificatory argument takes account of another piece of evidence by revising the probability distribution from the prior step so that it is logically possible for the prior step's intermediate conclusions and the new step's intermediate conclusions both to be perfectly calibrated. The agent sees herself as in each step taking account of some additional piece of evidence so as to preserve in her new intermediate conclusions the perfect calibration that began with her most prior probability distribution. She justifies her current opinion by showing herself entitled to take her current opinion to be perfectly calibrated–by showing how her current opinion results from calibration-preserving updates going back ultimately to her most prior opinions, whose calibration she is entitled to take as a working assumption.[33] I think that this picture explains why such an argument possesses the power to justify our current opinion regarding a hypothesis. For if we have shown ourselves entitled to believe a probability distribution to be perfectly calibrated, then (I argued) we have shown ourselves entitled to believe those degrees of confidence to be *right*, and so we have shown ourselves entitled to hold those degrees of confidence. When our 1905 physicist decided that he had previously failed to assign certain hypotheses the degrees of confidence they had merited considering the success of Planck's black-body equation, he was deciding that those earlier opinions had not been perfectly calibrated.

Note that I have argued neither that our current opinions nor that the intermediate conclusions in a justificatory argument must *be* perfectly calibrated in order for them to be justified. I have argued only that we must *justly believe* our current opinions to be perfectly calibrated in order for us to be justified in holding them, and that we justify so believing by showing how those opinions result from calibration-preserving updates, ultimately of a prior distribution that we are automatically entitled to take as calibrated. For us to be justified in believing certain opinions to be calibrated, it is not necessary that they actually be calibrated, but (given logical omniscience) their calibration must be logically possible.

Perfect calibration is nearly but not quite the sense of rightness appropriate for the intermediate conclusions at a given step of a justificatory argument. Suppose that at a step where '$\text{pr}(h) = .8$' is an intermediate conclusion, we discuss only one other hypothesis that we deem to be relevantly similar to h. It is logically impossible for exactly 80% of these two hypotheses to be true (since 80% of 2 is not an integer). So our interme-

[32] For any opinions consistent with our evidence, there are surely *some* prior probabilities that would yield these opinions via Bayesian updating on our evidence. But there is no guarantee that these prior probabilities would be "unbiased."

[33] In my examples, the agent shows this by presenting a *sequence* of updates, though she could in principle do so by a *single* update on all of her evidence.

diate conclusions at this step cannot be (justly believed to be) perfectly calibrated. Clearly, this is not from any shortcoming in those opinions; it is merely a technicality. To elaborate the "rightness" of these intermediate conclusions, we need what van Fraassen[34] calls "potential perfect calibration." Consider again the weather forecaster. Each day she assigns the next day to a certain reference class based on the current barometric pressure, temperature, and so forth. She judges the cases belonging to the same reference class to be "relevantly similar," that is, similar in the respects on which she bases her opinion regarding whether it will rain the next day. Accordingly, she assigns the same likelihood r of rain the next day to each of the co-classified cases. Now take the actual cases (days) belonging to a given reference class, some fraction f of which are rainy days. Consider what would happen to f were hypothetical further cases, relevantly like the actual cases, to be added. Consider what the limiting relative frequency would then be of rainy days among days where the forecaster's meteorological data fall within the relevant range. The forecaster's degrees of confidence are "potentially perfectly calibrated" exactly when for each reference class, f would approach arbitrarily close to r were the number of cases in the reference class to increase without bound. In an analogous way, we can specify what it is for the intermediate conclusions at a given step of a justificatory argument to be "potentially perfectly calibrated." They are potentially perfectly calibrated if and only if for any hypothesis h and number r, if '$\mathrm{pr}(h) = r$' is an intermediate conclusion at that step, then among hypotheses deemed at that step relevantly like h and accorded a degree of confidence (also r) at that step, the fraction of truths would approach arbitrarily close to r were there sufficiently many such hypotheses.[35]

As van Fraassen[36] explains, it is logically possible for one's opinions at a given time to be potentially perfectly calibrated only if those opinions satisfy the probability calculus. To glimpse the idea behind his argument, notice that, if we assign greater than 100% confidence to h, then it is logi-

[34] "Calibration: A Frequency Justification," pp. 304-05; "Belief and the Will," this JOURNAL, LXXXI, 5 (May 1984): 235-56, here p. 245.

[35] van Fraassen's concept of the limiting relative frequency of truths among relevantly similar hypotheses raises a host of familiar worries (such as how hypotheses are to be individuated, how the counterfactual 'Had there been more relevantly similar cases' is to be understood, whether the limiting relative frequency depends on the order in which the relevantly similar cases are considered, and so on)—familiar since analogous worries arise for certain frequentist views of objective probability. I shall not investigate here whether these worries apply to the use I wish to make of potential perfect calibration, or how they might be addressed.

[36] "Calibration: A Frequency Approach"; "Belief and the Will," pp. 245-46. See also Shimony, "An Adamite Derivation of the Calculus of Probability," in J.H. Fetzer, ed., *Probability and Causality* (Dordrecht: Reidel, 1988), pp. 151-61.

cally impossible, no matter how many hypotheses we are asked about that we deem relevantly like h, for more than 100% of these hypotheses to be true. Likewise, if we assign 60% confidence to h and 60% confidence to $\neg h$, then again, it is logically impossible, no matter how many hypotheses we are asked about that we deem relevantly like h (and so whose negations we deem relevantly like $\neg h$), for the fraction of truths among the hypotheses like h to approach 60% and the fraction of truths among the hypotheses like $\neg h$ to approach 60%, since that outcome would require that certain hypotheses and their negations both be true.[37]

I shall apply van Fraassen's argument to our opinions at a given step in a justificatory argument rather than to our opinions at a given moment. The argument then shows that it is logically possible for the intermediate conclusions at a given step of a justificatory argument to be potentially perfectly calibrated only if they satisfy the probability calculus. This constitutes a synchronic calibration argument. I shall now give a diachronic argument–again, not in the sense of applying to a shift in opinion from one moment to the next, but to a shift in opinion from one step in a justificatory argument to the next, as an additional piece of evidence is taken into account. Suppose $\mathrm{pr}(\cdot)$ is our probability distribution at a given step in a justificatory argument, and suppose $\mathrm{pr}'(\cdot)$ is our probability distribution at the next step in that argument, where e is the additional evidence that is brought to bear. Suppose $\mathrm{pr}(\cdot)$ is potentially perfectly calibrated. I now show that it is logically impossible for $\mathrm{pr}'(\cdot)$ and $\mathrm{pr}(\cdot)$ both to be potentially perfectly calibrated unless $\mathrm{pr}'(\cdot)$ results from updating $\mathrm{pr}(\cdot)$ according to BC.

[37] van Fraassen guarantees that the negation of a claim in h's reference class is in $\neg h$'s reference class: the reference class for 'x *is* $\neg A$' consists of claims of the form '... *is* $\neg A$', and if y is such that 'y *is* A' is in the reference class for 'x *is* A', then for every property B, our $\mathrm{pr}(x \textit{ is } B) = \mathrm{pr}(y \textit{ is } B)$, and so the reference class for 'x *is* $\neg A$' includes 'y *is* $\neg A$'. Does the scheme I sketched in footnote 31 require this sort of coordination between the reference classes for h and j where $j = \neg h$? It presumably would, once some account was added of what it means for an opinion to be "based" on a given consideration. But this coordination would be required only when we know all of the relevant logical truths, since only then could our reasons for our confidence in j be our reasons for our confidence in h along with the fact that $j = \neg h$. (In other words, if 'Γ possesses α' distinguishes h's reference class, then 'Γ possesses β' can distinguish j's, where β is the property of being logically equivalent to the negation of a claim possessing α.) I shall shortly discuss this point further in the main text.

In the argument that the logical possibility of potential perfect calibration requires coherence, logical omniscience is presumed (in order for various claims' reference classes to be "coordinated" as above). This seems reasonable, since in the absence of logical omniscience, it is unclear that coherence is required. I assume logical omniscience throughout. But it is interesting to note that on my conception of the reference classes, an agent could be calibrated even if she is ignorant of some relevant logical truths. She might, for instance, assign various number-theoretic hypotheses to various reference classes, depending on the reasons behind the degrees of confidence in $\langle 0,1\rangle$ that she currently assigns them. There is room here to treat the confirmation of logical truths.

Suppose that $\mathrm{pr}(\cdot)$ is potentially perfectly calibrated, so $\mathrm{pr}(h)$ equals the limiting relative frequency of truths among the hypotheses like h in the respects relevant to us at the given step. Roughly put, $\mathrm{pr}(h)$ is the value approached by the fraction of truths in a large set S of hypotheses each of which is like h in the respects relevant to us at this step. For $\mathrm{pr}'(\cdot)$ to be potentially perfectly calibrated, $\mathrm{pr}'(h)$ must equal the value approached by the fraction of truths in a large set of hypotheses each of which is like h in the respects relevant to us at the next step in the justificatory argument. But these respects are exactly the respects that were relevant to us at the previous step plus which e holds; the hypotheses like h in the respects relevant to us at step $(n+1)$ are exactly the hypotheses for which evidence like e holds and which are like h in the respects relevant to us at step (n). So for $\mathrm{pr}'(\cdot)$ to be calibrated, $\mathrm{pr}'(h)$ must equal the value approached by the fraction of truths among those hypotheses in S for which evidence like e holds. Since $\mathrm{pr}(\cdot)$ is potentially perfectly calibrated, $\mathrm{pr}(h \,\&\, e)$ is the fraction, among hypotheses in S, of truths for which evidence like e holds, while $\mathrm{pr}(e)$ is the fraction, among hypotheses in S, of hypotheses for which evidence like e holds. That is,

$$\mathrm{pr}'(h) = \frac{\begin{array}{c}\text{\# of hypotheses in } S \text{ that are true}\\ \text{and for which there is evidence like } e\end{array}}{\text{\# of hypotheses in } S \text{ for which there is evidence like } e}$$

$$\mathrm{pr}(e) = \frac{\text{\# of hypotheses in } S \text{ for which there is evidence like } e}{\text{\# of hypotheses in } S}$$

$$\mathrm{pr}(h \,\&\, e) = \frac{\begin{array}{c}\text{\# of hypotheses in } S \text{ that are true}\\ \text{and for which there is evidence like } e\end{array}}{\text{\# of hypotheses in } S}$$

in the limit of an arbitrarily large S. Hence, $\mathrm{pr}(h \,\&\, e)/\mathrm{pr}(e) = \mathrm{pr}'(h)$. But by definition, $\mathrm{pr}(h \,\&\, e)/\mathrm{pr}(e) = \mathrm{pr}(h|e)$. Therefore, $\mathrm{pr}'(h) = \mathrm{pr}(h|e)$. Thus it is shown that a justificatory argument must proceed by BC.

This reasoning nearly convinces me. The main source of my hesitation is that this argument depends on e's reference class "meshing" with h's. That is, since $\mathrm{pr}(\cdot)$ is potentially perfectly calibrated, $\mathrm{pr}(e)$ is the limiting relative frequency of truths among claims relevantly like e. Is part of what makes a claim relevantly like e that there is a corresponding hypothesis relevantly like h? The above argument presumes so, since it takes $\mathrm{pr}(e)$ to be the fraction, among hypotheses in S, of hypotheses for which there is evidence like e. But perhaps this presupposition is implausible.

Is some claim made relevantly like *e* partly by its relation to a hypothesis like *h* in the same way as two days are rendered relevantly similar, for the weather forecaster, partly by their each following days whose barometric readings are relevantly alike? Could a claim be relevantly like *e* but correspond to no hypothesis like *h*, and even if it did correspond to such a hypothesis, could this fact play no part in fixing its reference class?

But perhaps this "meshing" of *h*'s and *e*'s reference classes is exactly right. In the above argument, *h*'s reference class in the step where *e* is taken into account seems clearly to consist of those hypotheses in *S* for which evidence like *e* holds. Just as the previous day's barometric reading helps to fix the reference class to which a given day belongs, so in the later step of the justificatory argument, *h*'s being associated with evidence like *e* helps to fix *h*'s reference class. As *e* is bound up in *h*'s reference class, why should *h* not be bound up in *e*'s? Just as the weather forecaster determines the likelihood of rain on a given day by assigning that day to a reference class of relevantly similar days that she has experienced, so the scientist might determine the likelihood of making observation *e* by noting that *e* is strongly suggested by hypothesis *h* and thereby assigning *e* to a reference class of claims that are all strongly suggested by hypotheses like *h*. Notice that the same coordination of reference classes was required by the argument given earlier that, if we assign 60% confidence to *h* and 60% confidence to $\neg h$, then it is not logically possible for us to be potentially perfectly calibrated. That argument required that the hypotheses relevantly similar to $\neg h$ be the negations of the hypotheses relevantly similar to *h*. This seems correct.

If the reference classes must mesh in this way, then I have shown that the steps in a justificatory argument must conform to BC. In this role, I have argued, BC can be plugged into the familiar Bayesian proposals for explicating various features of confirmation in science. Notice, again, that no analogous argument can be made that scientists must accord with BC in updating their opinions from one moment to the next, upon receiving new evidence. At any moment, I have argued, a scientist must believe her *current* opinions (and the intermediate conclusions in the arguments by which she justifies them) to be potentially perfectly calibrated, but she may believe that her opinions *at other moments* fail(ed) to be potentially perfectly calibrated. She may change her mind–at one time believing, at another moment denying that certain opinions match the objective frequencies.[38]

[38] Since she need not believe her past opinions to have been calibrated, the connection I have demonstrated between calibration and BC does not show that, upon receiving new evidence, she must diachronically update her opinions in accordance with BC. It might be argued that an agent must not only *believe* her opinions to be calibrated, but for those opinions genuinely to be justified, they must *be* calibrated. It would follow that

IV

As I noted, a diachronic Dutch Book argument purports to show that BC is the only acceptable *rule* for updating opinion. It thus leaves room for van Fraassen's position: that we can violate BC without vulnerability to a Dutch Book if we use no rule for updating our opinions. But the above calibration argument leaves no room for this position. It shows that the intermediate conclusions in a justificatory argument can logically possibly all be potentially perfectly calibrated *only if these conclusions succeed each other according to BC.* No reference is made to whether some rule is consulted in proceeding from one step to the next.

It might be objected that my argument *does* presuppose that some rule is involved–namely, a rule fixing which hypotheses are (by the agent's own lights) relevantly similar (at a given step) to *h*. If the agent endorses no such rule, that is, recognizes no sense of hypotheses "relevantly similar" to *h*, then (the objection runs) there is no possibility of her reference classes being such that her opinions cannot logically possibly be potentially perfectly calibrated–because she has no reference classes.

But my suggestion was *not* that, *if* the agent recognizes a sense of "relevant similarity" to *h*, then she must believe that her intermediate conclusion regarding *h* could logically possibly be potentially perfectly calibrated, and so the transitions between the steps of a justificatory argument must accord with BC. Rather, I suggested that a rational agent *must* recognize a sense of "relevant similarity" to *h*. That is because she must believe that her opinion of *h* is potentially perfectly calibrated, since this is to believe it to be *right* (considering the evidence brought to bear by that step). She must believe it to be right since she believes that it should guide her in taking account of some further evidence when moving to the next step in the justificatory argument.

It is worth noting another difference between the role in the Dutch Book argument played by a rule for updating and the role in my ar-

when an agent holds justified prior opinions and then, upon acquiring new evidence, adopts justified new opinions, this updating must conform to BC. But my argument that an agent must believe her opinions at a given step of a justification to be calibrated was, in brief, that otherwise she would have no reason to base her opinions at the next step on the opinions at the previous step. This argument does not somehow generalize to show that an agent's opinions must really be calibrated. The opinions at one step of a justificatory argument are based on the opinions at the previous step, but the same does not hold diachronically; an agent need not be guided by her old opinions in forming new ones. That is why I am willing to require that an agent *believe* her opinions to be calibrated, but not to require that agent's opinions really be calibrated.

Whether justified opinions must *be* calibrated turns on issues concerning epistemological externalism that are beyond this paper's scope. Of course, any requirement that justified opinions be calibrated would remain crucially subjective, considering that the reference classes depend on which hypotheses *the agent considers* relevantly similar.

gument played by a rule determining which hypotheses are relevantly similar (at a given step) to h. The Dutch Book defense of BC requires that there be some particular rule for updating that the agent has explicitly adopted (so the bookmaker also knows in advance how the agent would update, and can design the book accordingly). In contrast, my argument does not presume the agent to have explicitly committed herself to some particular rule for determining which hypotheses are relevantly similar (at a given step) to h. Furthermore, the Dutch Book argument presumes that the agent has a disposition to update her opinions in a certain way if she receives certain evidence (and that the agent knows what her disposition is, since the bookmaker must know). In contrast, the calibration argument does not presume that the agent currently has some disposition to believe, under certain circumstances, that a given hypothesis is relevantly similar to h. The agent may have no such disposition because she does not recognize much about the sense of relevant similarity upon which she is relying. That is, she may not have identified precisely what her reasons are for her opinion (at a given step) regarding h, so she may have no disposition to believe that her reasons in another case, for the opinion she would hold there regarding some hypothesis, would be the same as her reasons (at this step) for her opinion regarding h.[39] The calibration argument presumes only that the agent *has* such reasons–that her opinions are grounded in some sense of "relevant similarity." This suffices, I have argued, to require that her justificatory arguments conform to BC.

van Fraassen defends a principle he calls *reflection*: that at any moment t, x is your degree of confidence that h is true in the event that at t or some time thereafter, x is your degree of confidence in h. (That is, reflection says that for $t \leq t'$, $\mathrm{pr}_t\ (h|\mathrm{pr}_{t'}(h)=x)\ =\ x$.) Roughly speaking, reflection expresses an agent's belief that now and in the future, her opinions are calibrated.[40] van Fraassen[41] has shown

[39] To *have* reasons, it is not necessary that you be in a position to articulate them, or even to say whether your reasons in one case are the same as your reasons in another. To do that may require careful philosophical thought. (For more on what it takes for an opinion to be *based* on a given consideration, see Lehrer (*op. cit.*) and Pappas (*op. cit.*) and the references therein.) The agent might acquire a disposition to believe that (at a given step) she has the same reasons for her opinions regarding two hypotheses. Such a disposition might be groomed by asking the agent to address questions like 'Why do you have a different degree of confidence in Brown's five-year survival after a heart transplant than Smith's, considering that they are similar in all of these respects...?'

[40] For a careful discussion of what reflection requires, see Mitchell Green and Christopher Hitchcock, "Reflections on Reflection," *Synthese*, XCVIII (1994): 297-324.

[41] "Belief and the Will."

that an agent who updates her opinions from one moment to the next according to BC must conform to reflection; Phillip Dawid[42] has shown likewise that a coherent Bayesian forecaster must assign probability one to the event that she is and will be potentially perfectly calibrated. van Fraassen and Dawid then ask how such confidence in one's own current and future calibration could be warranted.

What I showed in the previous section is not van Fraassen's or Dawid's result, but more like their converse. I showed that an agent who is potentially perfectly calibrated before and after an update must have updated in accordance with BC. This result is not quite the converse of van Fraassen's or Dawid's because it concerns whether an agent *is* potentially perfectly calibrated, not whether she *believes* herself so to be. Of course, I have also argued that an agent must believe in the potential perfect calibration of the intermediate conclusions at each step of her justificatory arguments. In particular, then, she must believe that the opinions at the final step of her justificatory argument (that is, her *current* degrees of belief) are potentially perfectly calibrated. But I have not argued that she must believe herself to be calibrated at each *future* moment. On the contrary, such a requirement seems just as counterintuitive as requiring that she believe herself to have been calibrated at each past moment.[43] I have emphasized that an agent who at one time believes certain opinions to be merited by certain evidence may later change her mind. Although BC applied to diachronic updatings entails reflection, I avoid requiring that an agent believe in the calibration of her future selves, since I fail to require an agent to conform to BC in updating her opinion from one moment to the next, as new evidence is received. Instead, I regard BC as governing the steps of her justificatory arguments.

The philosophical challenges here lie not so much in giving the formal argument, at the close of the previous section, as in understanding just what that argument shows. Hence, the bulk of this paper was devoted to arguing (i) that an agent must believe the intermediate conclusions in her justificatory arguments to be potentially perfectly calibrated, and (ii) that BC could avoid certain important difficulties if it was applied within justificatory arguments rather than diachronically. I hope to have suggested a new way that BC might account for various features of theory confirmation in science.

MARK LANGE

University of Washington

[42] "The Well-Calibrated Bayesian" (with discussion), *Journal of the American Statistical Association*, LXXVII (1982): 605-13.

[43] van Fraassen addresses objections to reflection in his "Belief and the Problem of Ulysses and the Sirens."