Inference to the Best Explanation, Bayesianism, and Knowledge

Alexander Bird

Abstract

How do we reconcile the claim of Bayesianism to be a correct normative theory of scientific reasoning with the explanationists' claim that Inference to the Best Explanation provides a correct description of our inferential practices? I first develop Peter Lipton's approach to compatibilism and then argue that three challenges remain, focusing in particular on the idea that IBE can lead to knowledge. Answering those challenges requires renouncing standard Bayesianism's commitment to personalism, while also going beyond objective Bayesianism regarding the constraints on good priors. The result is a non-standard, super-objective Bayesianism that identifies probabilities with evaluations of plausibility in the light of the evidence, conforming to Williamson's account of evidential probability.
Keywords explanationism, IBE, Bayesianism, evidential probability, knowledge, plausibility, credences.

1 Introduction

How do scientists reason? How ought they reason? These two questions, the first descriptive and the second normative, are very different. However, if one thinks, as a scientific realist might, that scientists on the whole reason as they ought to then one will expect the two questions to have very similar answers.

That fact bears on debates between Bayesians and explanationists. When we evaluate a scientific hypothesis, we may find that the evidence supports the hypothesis to some degree, but not to such a degree that we would be willing to assert outright that the hypothesis is true. Bayesianism is a theory about how such evaluations should be made. *Standard Bayesianism* I take to be the view that involves the following commitments:

> *Personalism* A thinker's evaluations in such cases are *credences*, doxastic states of the subject that may be operationalized as dispositions to accept a set of bets on the truth of propositions;
> *Probabilism* The evaluations, if rational, can be represented as *probabilities*; they should obey the axioms of probability;
> *Conditionalization* The evaluations are subject to *updating* in the light of new evidence e, so that the new evaluation of a hypothesis $P_n(h)$, should be equal to the old evaluation of the hypothesis conditional on that evidence $P_o(h|e)$.

The standard Bayesian holds that credences are an analogue to belief simpliciter: traditional epistemology concerns what it is rational to believe outright while

Bayesian epistemology concerns the rationality of partial belief. Various arguments (Dutch book arguments, accuracy-dominance arguments) are used to show that a personalist should be a probabilist, i.e. that rational credences are probabilities.[1] If probabilism is true, then the evaluation function, P, is also a probability distribution. And a consequence of the probability axioms, Bayes's theorem, holds of the evaluations of any propositions, e, h, and their conditional relations:

$$\text{(B)} \qquad P(h|e) = \frac{P(e|h)P(h)}{P(e)}.$$

Conditionalization says that we should update those evaluations (probability distributions) in the light of new evidence; so if e is in fact new evidence, then:

$$\text{(S)} \qquad P_n(h) = P_o(h|e) = \frac{P_o(e|h)P_o(h)}{P_o(e)},$$

where P_o is the probability distribution before the evidence was acquired and P_n is the new probability distribution in the light of that evidence. In the context of conditionalization, $P_o(h)$ is the prior probability of the hypothesis, $P_o(e)$ is the prior probability of the evidence, also known as the expectedness of the evidence, and $P_o(e|h)$ is the likelihood; $P_o(h)$ is the posterior probability, the probability now given the hypothesis having received the evidence. (S) is known as *strict* conditionalization. As we shall see, strict conditionalization is not the only updating rule that the Bayesian may employ.

Standard Bayesianism is normative, a view that asserts requirements about how reasoning ought to proceed, in addition to a descriptive kernel (personalism). A standard Bayesian is also a subjectivist if she holds that probabilism and conditionalization are the only rational constraints on credences. An objective Bayesian extends standard Bayesianism with further constraints, such as a principle of indifference.[2] Some Bayesians are non-standard in that they do not accept all the features of standard Bayesianism. For example, one might accept that there are counterexamples to Conditionalization.

Descriptive explanationism I take to be the view that people, scientists included, often hold the explanatory character of a hypothesis to be relevant to its epistemic evaluation. Inference to the Best Explanation (IBE), understood as a description of an inferential practice central to science, is explanationism *par excellence*: it holds that such subjects come to accept a hypothesis because it provides a better explanation of the evidence than its rivals. Per se, normative and descriptive claims are easy to reconcile. Even if IBE and Bayesianism are entirely different, it might be that IBE describes how we do reason while Bayesianism describes how we ought to reason (but do not).

A scientific realist may well claim that scientists do reason, by and large, as they ought to: IBE is both descriptively and normatively correct. In which case the task of reconciliation becomes pressing. Some anti-realists may escape the issue by denying the descriptive or normative claims of IBE or both. Notable critics of IBE, such as van Fraassen (1989:160–70) and Salmon (2001), think that they are not compatible.

[1] See Joyce (1998); Leitgeb and Pettigrew (2010a,b). The argument that Bayesian conditionalization, in addition to Bayesian coherence (the probabilist requirement that a subject's credences satisfy the probability axioms), is required by rationality is thus a *diachronic* Dutch book argument. Whether a diachronic Dutch book implies irrationality is far from clear (Briggs 2009; Mahtani 2012).

[2] Other extensions to standard Bayesianism include commitments to Regularity, Countable Additivity, the Principal Principle, and Maximum Entropy.

Van Fraassen argues that if explanationist considerations were relevant, then they would encourage a thinker to boost the credence $p(h|e)$ above that mandated by Bayes's theorem, when h is a better explanation than competitor hypotheses while reducing the corresponding credences for those competitors. Such a thinker would be open to a Dutch book (a set of bets he takes to be fair but which will lose money whatever the outcome). An account of reasoning cannot be normatively correct if it enjoins accepting a Dutch book. So IBE is not a normatively correct account of reasoning.

So the explanationist who holds that IBE is the right answer to both the descriptive and normative questions must reject van Fraassen's claim that IBE would encourage different credences from those that issue from Bayes's theorem. Rather, Bayesianism and IBE are compatible answers to both questions. Since Bayesianism and IBE are prima facie so very different, showing that they are compatible is not straightforward.

In this paper I develop three challenges for compatibilism between Bayesianism and explanationism. These three challenges relate to the fact that IBE ought on some occasions to be able to give us *knowledge*. To answer the challenges I argue that we should (i) reject personalism, replacing credences by assessments of plausibility (hence departing from standard Bayesianism), and (ii) embrace eliminative abduction.

2 Heuristic compatibilism

Peter Lipton (2004:103–20) has argued that there is in fact no incompatibility between Bayesianism and IBE.[3] In Lipton's view IBE is a *heuristic* we employ in lieu of directly applying (S). This heuristic view of compatibilism is examined and advanced in detail also by Ted Poston (2014:149–81).[4] In this section I articulate how I think this idea should be understood.

A kind of problem may have an optimal method of solution, but that method may not be easy to implement properly. Implementing it properly might be costly in terms of effort. Or that attempt may be prone to error, and so lead to a mistaken answer. A heuristic for solving a problem is a method that may be in principle suboptimal but is more easily implemented. It may well be rational to employ the heuristic if the saving of effort outweighs the loss in reliability. Or the heuristic might be more reliable than the optimal method as poorly implemented by us inadequate thinkers. For example, the recognition heuristic is employed when subjects are asked to say which of two named cities has a greater population: when only one city is recognized, then judge that that city has the larger population. This is both easy to implement in the relevant cases; and it turns out also to be more reliable than the process of making inferences from relevant knowledge (subjects performed *less* well when they knew something about both cities)—this is the *less is more* effect.

Lipton notes the considerable evidence that we find thinking with probabilities difficult and are prone to error. So it would be rational for us to employ a heuristic in place of Bayesian reasoning. I propose that the heuristic is this:

[3] Lipton does not say which form of Bayesianism he has in mind, so I assume that his focus concerns standard Bayesianism in its subjectivist form.

[4] See also Schupbach (2016) in this volume for an account of explanatory power that has a heuristic value relative to Bayesianism.

(H) If hypothesis A is a lovelier potential explanation of the evidence than B, then judge A to be more likely to be true than B.

('Loveliness' is Lipton's term for the explanatory goodness of a potential explanation; it is what, according to explanationism, we evaluate when evaluating a hypothesis.) If (H) is a rational heuristic for Bayesianism then it should yield much the same results as (S) would. We can show that it does by arguing that elements of the explanationist heuristic correspond to components of the Bayesian updating in (S). The can be done, in Lipton's view, as follows. (I) If propositions h and e are ones that the subject has not given thought to before, then considering both as hypotheses, we make as assessment of their explanatory loveliness, taking into account such features as unification and simplicity.[5] This corresponds to the Bayesian process of assigning values to the priors $P_o(h)$ and $P_o(e)$. (II) We evaluate the explanatory loveliness of h in relation to the evidence e, and we then change the overall loveliness assessment of h accordingly. This corresponds to conditionalization, (S), whereby we move from the prior probabilities, $P_o(h)$ and $P_o(e)$ to the posterior probability, $P_n(h)$. (III) Explanationist considerations determine which evidence is relevant to the hypothesis—the injunction to consider the total evidence, although implicit in Bayesianism, is not one we can actually implement, so we need a heuristic to guide us to the relevant evidence, viz. whether the evidence is explained by the hypothesis.

To complete the explanation of why IBE may be understood as a heuristic for Bayesianism, we should show that using IBE will tend to get one similar results to Bayesianism. For example, we should show that (H) says the we should judge that A is more likely than B when (S) says that P(A) is greater than P(B). First, however, I shall expand on the details of Lipton's proposal a little. Consider the question: how good an explanation is hypothesis h? Now some aspects of the answer will concern h alone, whereas others will concern h's explanatory relationship to the evidence e. For example the simplicity of h will concern h alone. Newton's law of gravitation plus the laws of motion have a high degree of simplicity and strength, and that fact (other things being equal) makes it a lovelier theory than one with more laws or more complex ones. On the other hand, whether the explanation of e by h is lovely is a matter of the relationship between the two. For example, the hypothesis that an asteroid impact caused the K-T (K-Pg) extinction is a lovely explanation of the fact that the geological K-T layer coincides with an iridium anomaly, because the fact that asteroids have high iridium concentrations directly explains that coincidence. Following Einstein, I call these the *internal* and *external* explanatory virtues of a hypothesis.[6]

[5] Richard Pettigrew has pointed out a potential pitfall for compatibilism. In some case h itself determines a probability for e. Could it be that two competing hypotheses, h_1 and h_2 both give the same value to the probability of e yet h_1 is a better explanation of e then h_2? If so, then should that fact be reflected, according to the explanationist, in a higher value for $P(e|h_1)$ than for $P(e|h_2)$? If it were, then that would conflict with the Principal Principle, which says, in effect, that $P(e|h)$ must be equal to the probability that h determines for e. (And it would would also amount to another instance of van Fraassen's kind of criticism.) I think that the compatibilist must regard the probabilities $P(e|h_1)$ and $P(e|h_2)$ as capturing all that there is to be said about explanatory goodness. In effect, explanatory goodness has a role to play when h does not entail a value for e.

[6] In his "Autobiographical Notes" Einstein (1991) contrasted:

> Lipton himself needs The first point of view ... the theory must not contradict empirical facts ... [it] is concerned with the confirmation of the theoretical foundation by the available empirical facts. The second point of view is not concerned with the relation of the material of observation but with the premises of the theory itself, with what may briefly but vaguely be characterized as the 'naturalness' or 'logical simplicity' of the premises

Armed with this distinction between the internal virtues of h per se and the external virtues of h's explanation of e, we can articulate a slightly different set of correspondences between IBE and Bayesianism and argue that they are plausible.

First, the internal explanatory virtues of h correspond to $P(h)$, insofar as h is a novel hypothesis. Subjective Bayesianism specifies no constraints on the priors, and so there is no danger of an explanationist account of the internal explanatory virtues of h conflicting with the requirements of the subjective Bayesian. As far as the latter is concerned the priors can be anything—they could come from an ouija board, so they might as well come from explanationist considerations. However, the heuristic view aims at showing that the heuristic (H) (mostly) yields the conclusion that an ideal Bayesian would reach. And that would not be the case if the Bayesian were to choose unusual priors (which the subjective Bayesian will feel entitled to do). Nonetheless, as a matter of descriptive fact, our actual assessment of the probability of priors does correlate with our assessment of their explanatory value. In advance of the evidence we do think that simpler, more unifying explanations are more likely to be true.

Indeed, the very propositions that we regard as prime drivers of updating reflect this. Consider the principle of indifference—one should accord equal probability to propositions regarding which one's evidence is symmetric (i.e. one's evidence bears the same relationship to each outcome; see Jon Williamson's (2010:2) 'equivocation norm'). For example, nothing that one knows about a coin discriminates between heads and tails: it might be fair; or it might be biased, but if it is biased one knows nothing that says that it would be biased towards heads rather than tails or vice-versa. If so, one should accord the hypothesis that the next toss will yield tails the same probability as the hypothesis that it will land heads. Now think about a sequence of ten tosses that one is about to make. Here are two approaches to thinking about the outcome by applying the principle of indifference. Approach 1: consider hypotheses each of which specifies how the sequence of tosses will turn out—there are 1,024 such hypotheses. Apply the principle of indifference to these, giving each a probability of 1/1,024. One can then calculate the probability of various proportions of heads/tails in the ten tosses. Approach 2: consider first the hypotheses that assert the proportion of heads/tails—there are eleven such hypotheses. Apply the principle of indifference to these eleven hypotheses of the second type, giving each a probability of 1/11. One can then use these probabilities to calculate probabilities for particular sequences. The two approaches lead to different probabilities for particular sequences: the first option says that a sequence of 10 heads has a probability of 0.001 whereas the second option says that that sequence has a probability of 0.09. Likewise the two approaches lead to different probabilities for the proportions of heads/tails. More importantly, Huemer (2009) shows that the choice leads to a difference between scepticism and inductivism: the first choice is sceptical, in that it does not allow the probability assigned to heads on the tenth toss to change, even if one has learned that the first nine tosses have all yielded heads, whereas the second choice does assign a high probability to heads in the light of such evidence. The second set of hypotheses are explanatory hypotheses whereas the first set are not. And that, argue Huemer (2009) and Poston (2014), shows that Bayesianism *needs*

> ...The second point of view may briefly be characterized as concerning itself with the 'inner perfection' of a theory, whereas the first point of view refers to the 'external confirmation'.

Quoted by Salmon (1990:181).

explanationism if Bayesianism is to be a satisfactory account of inductive reasoning. Descriptively, we do think that we should consider the hypotheses concerning proportions/probabilities first and allow the probabilities concerning particular sequences to follow as a consequence rather than vice-versa. That strongly suggests that we are indeed moved by explanatory considerations in the assignment of priors to hypotheses. Put another way, we are as a matter of fact inductive thinkers, and since being a Bayesian inductivist requires (or at least motivates) explanationist thinking about priors, then we should expect that we are explanationists about priors.

Secondly, the external explanatory virtues of h relative to e correspond to the *ratio* of the likelihood to the expectedness of the evidence, $P(e|h)/P(e)$. External explanationist considerations do not correspond to the likelihood $P(e|h)$ alone. As Lipton appreciates, $P(e|h)$ can be very high even if h does *not* explain e well—if e is itself very probable: famously Jim's taking a contraceptive pill does not provide a good explanation of his not getting pregnant, since it was certain anyway that he would not get pregnant; still $P(e|h)$ is very high. And, in reverse, $P(e|h)$ can be low even if h *does* explain e well. For example, Holmes suspects Professor Moriarty of having murdered Major Arbuthnot and of having disposed of his body, meticulously erasing evidence of the crime. Holmes then discovers a minute trace of the highly unusual Uzbek tobacco smoked by Arbuthnot in the cuff of Moriarty's coat. Now given the meticulousness that the hypothesis ascribes to Moriarty, finding that tobacco trace is itself highly unexpected, so $P(e|h)$ is low. But Holmes's hypothesis is still an excellent explanation. The latter fact is still consistent with the demands of Bayes's theorem, since $P(e)$ is even lower. I propose therefore that the external explanatory virtue of h in connection to e corresponds to the *ratio* $P(e|h)/P(e)$, i.e. the measure of how much more probable e becomes when conditioned on h.

Bayes's theorem states that the ratio $P(e|h)/P(e)$ is equal to $P(h|e)/P(h)$, the ratio of the probability of the hypothesis conditional on the evidence to the prior probability of the hypothesis. So, given conditionalization, $P(e|h)/P(e)$ is equal to the proportion by which the probability of the hypothesis changes on learning the evidence. $P(h|e)/P(h)$, or some monotonically strictly increasing function thereof (such as its logarithm), is taken by some Bayesians to be a measure of the confirmation of h by e (Keynes 1921; Horwich 1982; Milne 1996). And so the proposal that (external) explanationist considerations correspond to $P(e|h)/P(e)$ and so correlate with this measure of confirmation means that an explanationist account of confirmation corresponds to the Bayesian one.

The heuristic view holds that Bayesian reasoning and explanationist reasoning are distinct, although leading to the same or similar judgments reasonably frequently. We should therefore not expect every component in the Bayesian calculation to have an explanationist counterpart. So a subject forming a judgment on the basis of explanationist considerations forms a judgment that corresponds to a Bayesian value for $P(e|h)/P(e)$ without that being the result of preceding judgments corresponding to $P(e|h)$ and $P(e)$. That is an advantage of the view I propose, since it means that explanatory virtue can support a hypothesis even if we have no clear idea of what $P(e|h)$ and $P(e)$ should reasonably be: Holmes might have only an imprecise idea regarding the probability of the presence of Uzbek tobacco ash in Moriarty's cuff, both the unconditional probability and the probability conditional on Moriarty having carried out a carefully planned murder. Even so, he might be able to be rather more precise about the degree to which the Moriarty hypothesis *increases* the probability of the presence of the ash. For the heuristic view to hold,

it should be the case that as a matter of fact, subjects' judgments of the external explanatory virtues of a hypothesis, i.e. when the evidence confirms a hypothesis by explanationist lights, should correlate with their assessments that the hypothesis make the evidence more likely than it would otherwise be. And that looks correct: as the Holmes example and others show, if a hypothesis explains some evidence, then we think that the evidence proposition is more likely if the hypothesis is true than if we don't make that assumption.

Having drawn correspondences between the explanatory virtues (internal and external) of a hypothesis and components of the Bayesian calculation, to confirm the heuristic view we need to show that our overall assessment of the probability of a hypothesis correlates with something that we can regard as the 'product' of our judgments regarding the internal explanatory virtue of a hypothesis and its external explanatory virtue. Since our judgments of explanatory virtue are not numerical, it is difficult to show this explicitly. On the other hand, the way we combine those judgments does seem to mirror multiplication rather than addition or averaging. For example, if a hypothesis is known to be false then even unlikely evidence it explains well will not make us think that the hypothesis is even moderately likely—that corresponds to the product of a finite amount and zero, whereas averaging would suggest that a known false hypothesis could become likely on favourable evidence.

To conclude: the challenge for compatibilism between explanationism and Bayesianism was that if the two give different results regarding the credence we should attach to a proposition or the ordering of our preferences among propositions, then we would have a reason for thinking that explanationism is incoherent. If, on the other hand, the two always agree, the explanationism is otiose: Bayesianism gives a precise and well-motivated account of confirmation, whereas explanationism is imprecise and philosophically controversial. The heuristic approach advocated by Lipton and Poston resolves this challenge, by showing that explanationism is a useful substitute for Bayesian reasoning. For various reasons it is unnatural and difficult for human thinkers to implement the probabilistic reasoning required by standard Bayesianism. So we use explanationist considerations as a heuristic—a manner of reasoning that will by and large get the same result as Bayesian reasoning but without its costs and pitfalls. I have added some detail to this picture, showing how aspects of explanationist thinking correspond to elements of Bayes's theorem. The internal explanatory virtues of a hypothesis correspond to $P(h)$; the external explanatory virtues correspond to $P(e|h)/P(e)$; and we combine these virtues in a manner that corresponds to their product. Explanationism does not give precise values for probabilities, but it does allow us to order our preferences among hypothesis. According to the heuristic approach, the ordering that the heuristic (H) is for the most part the same ordering that explicit and accurate use of Bayes's theorem and conditionalization would have given us had we been in a position to use it.

3 Three challenges for compatibilism

In this section I raise three challenges to the compatibilism just articulated. These start from the assumption that IBE can lead to knowledge. One might reject that assumption, but if explanationism is a correct descriptive account of some important inferential practices in science, then that rejection amounts to scepticism about those parts of science that employ those practices.

3.1 Knowledge and objectivity

What does this compatibilist argument really show? It shows that if Bayesianism is the correct normative account of scientific reasoning, then explanationism need not be regarded as in conflict with Bayesianism. On the contrary. explanationism describes the heuristic mechanism whereby scientists typically implement Bayesianism. While this ought to satisfy the subjective Bayesian, it ought not satisfy the explanationist. For the latter thinks that there is more to epistemology than is found in subjective Bayesianism. The explanationist thinks that on some occasions IBE can give us knowledge. She holds explanationist considerations to be a guide to the truth. So the explanationist cannot regard standard Bayesianism as a complete normative epistemology. She should say that priors and likelihoods informed by explanationism are better than those not so informed. Explanationist Bayesianism must reject subjective Bayesianism, and a strong form of objective Bayesianism looks more suitable, with constraints on which priors rationally acceptable.[7] Some priors—those guided by explanationist considerations—are better or more acceptable in that they will lead to a posterior probability function such that hypotheses with high probabilities will more often be true than those with low probabilities.

Lipton himself needs to say something like this. Otherwise many of the prima facie problems that he raises for IBE (and addresses) would not be concerns at all. For example, Voltaire's problem asks why it is that we should think that the explanatorily loveliest explanation is the actual (true) explanation. Should we think that the actual world is the loveliest of all possible worlds? But that's irrelevant if *all* the epistemic work is done by Bayes's theorem while explanationist concerns are merely the source of the values of the priors and likelihood, without making any normative contribution. However, concern with Voltaire's problem is required if we want the resulting epistemic attitude to correlate with truth and IBE possibly to lead to knowledge.

So while Lipton's compatibilism can defuse the criticism from subjective Bayesians, the explanationist holds that there is more to be said about the epistemological function of explanationist factors—specifically regarding their connection to truth and knowledge.

3.2 Knowledge, evidence, and credence

If a subject has already acquired evidence e, then $P(e)$ is trivially equally to 1. If b is a proposition known by the subject, then $P(b)$ also 1. This is a trivial consequence of the form of Bayes's theorem that takes background knowledge into consideration:

$$(\mathrm{B}') \qquad P(h|e\&b) = \frac{P(e|h\&b)P(h\&b)}{P(e\&b)}.$$

In effect Bayesianism treats evidence and background knowledge equivalently. (I take the significance of 'background' to be simply knowledge other than e, which is knowledge in the foreground because when conditionalizing on new evidence e it is the probability of e that changes to 1 whereas the probability of b is already at 1.) In a sense, therefore, Bayesianism endorses Timothy Williamson's equation of evidence and knowledge, E=K (Williamson 1997). (Alternatively, if one accepts E=K for Williamson's independent reasons, then that implies that the propositions that make up b in (B′) really are the propositions *known* to the subject rather than, e.g., those believed by the subject.)

[7] See Psillos (2004) and Huemer (2009) on explanationism as a rational constraint on priors.

Yet the view that $P(e) = 1$ and $P(b) = 1$ has highly unpalatable consequences for personalism (the view that the probabilities in question are credences, canonically measured by betting quotients). For a credence of 1 in p means that one would be willing to take a bet that would yield a one penny gain if p is true and unending torture if p is false. But few if any propositions are such that reasonable people would take such a bet. Certainly the contingent propositions that we take to be knowledge or to be our evidence are not typically of that sort. That looks to lead to the sceptical conclusion that we do not know those propositions after all and that no such propositions can ever really be our evidence. Furthermore, that conclusion, apart from being sceptical, also means that Bayesian updating (conditionalization) will never occur, thereby rendering it pointless.

3.2.1 Jeffrey conditionalization

This concern might to lead one away from standard ('strict') Bayesian conditionalization to Jeffrey conditionalization (Jeffrey 1983) (so long as one retains personalism—but see below). Jeffrey conditionalization does not conditionalize just on propositions that have probability equal to 1 (as in strict conditionalization), but also on propositions stating 'uncertain evidence'. If a proposition is uncertain, then so is its negation. Consequently, Jeffrey conditionalization conditionalizes both on the uncertain evidence proposition and on its negation, weighted according to one's subjective confidence. So:

$$\text{(J)} \qquad P_n(h) = \frac{P_o(e|h)P_o(h)}{P_o(e)}P_n(e) + \frac{P_o(\neg e|h)P_o(h)}{P_o(\neg e)}P_n(\neg e)$$

where $P(e)_n$ and $P(\neg e)_n$ are the weightings, *viz.* the new credences in e and its negation, resulting from some observation for example.

Strict conditionalization is the special case of Jeffrey conditionalization where $P(e)_n = 1$ (and so $P(\neg e)_n = 0$). So both strict conditionalization and Jeffrey conditionalization involve a process where a proposition, e, changes its credence in a way that is not itself the result of conditionalization, from $P(e)_o$ to $P(e)_n$, where $P(e)_n = 1$ for the strict conditionalization case and $P(e)_n < 1$ for the Jeffrey conditionalization case.

Note that both Jeffrey and strict conditionalization can be seen as special cases of a more general Bayesian updating rule that is a consequence of finite additivity, the third of the Kolmogorov probability axioms:

$$\text{(F)} \qquad P(a) = P(a \wedge b_1) + \ldots + P(a \wedge b_i)$$

where $\{b_1 \ldots b_i\}$ is a partition, i.e. a set of mutually exclusive and jointly exhaustive propositions. Given the definition of conditional probability, (F) is equivalent to:

$$\text{(F}') \qquad P(a) = P(a|b_1)P(b_1) + \ldots + P(a|b_i)P(b_i).$$

Let us say some evidence comes in that (rationally) causes the subject to change some of the $P(b_1) \ldots P(b_i)$, so they are now $P_n(b_1) \ldots P_n(b_i)$. Then the Bayesian will expect the subject to change $P(a)$ accordingly to $P_n(a)$:

$$\text{(F}'') \qquad P_n(a) = P_o(a|b_1)P_n(b_1) + \ldots + P_o(a|b_i)P_n(b_i)$$

(where P_o is the old probability distribution P in (F′)). The proposition a will itself be a member of another partition, so that $P(c)$ stands to $P(a)$ and $P(c|a)$ just as $P(a)$

stands to $P(b_1)$ and $P(a|b_1)$ in (F′). So a change in the $P(b_1)\ldots P(b_n)$ has a ripple effect, requiring the updating of the probabilities of many propositions to which those $b_1\ldots b_n$ are related, transitively, by equations such as (F″).

Now note that Bayes's theorem tells us that $P_o(a|b_1) = P_o(b_1|a)P_o(a)/P_o(b_1)$ etc. Hence:

$$(\mathrm{F}''') \qquad P_n(a) = \frac{P_o(b_1|a)P_o(b_1)}{P_o(b_1)}P_n(b_1) + \ldots + \frac{P_o(b_i|a)P_o(b_i)}{P_o(b_i)}P_n(b_i).$$

Since (F″) is an updating rule so is (F‴). An updating rule identical to Jeffrey conditionalization, (J), is the special case of (F‴) where we update using the partition $\{e, \neg e\}$.

So what is distinctive about Jeffrey conditionalization is not the updating rule, which is in use whenever there is a change in the probability distribution. For example, as a result of strict conditionalization, some probabilities will change and (F‴) and its kin will be used iteratively to update further propositions. So what is at issue is not the updating rule but instead which changes in the probability distribution should rationally *initiate* updating.

The strict conditionalizer has a clear if incomplete answer to that question. Updating is initiated when a proposition achieves the status of evidence. This answer is incomplete in two respects. First Bayesianism tells us almost nothing about when a proposition is evidence—the only constraint is that evidence propositions are those propositions that have a probability equal to 1. Secondly, for updating to be a rational requirement, the change of a proposition's status to evidence must itself be rational. If a subject who starts with a rational probability distribution changes the probability he attributes to a proposition p from less than 1 to 1 without any rational cause, then it is not rationally required that the subject update the rest of his probability distribution in line with strict conditionalization. Another option (and the one that does seem rationally required) is to reverse the change. Put another way, if the subject keeps the rest of his probability distribution fixed (which he ought, given that it was rational to begin with and the subject has undergone no rational change), then he is rationally obliged to change $P(p)$ again, in line with Bayes's theorem, which will set $P(p)$ back to where it started.

Consequently the Bayesian has to outsource some important epistemological questions: Which propositions are evidence propositions? and Why is it rational to update when a proposition becomes an evidence proposition?[8] As we shall see, a response can be given that answers both questions. Evidence propositions are precisely those that are known; as already discussed, known propositions are indeed accorded probability of 1. And clearly if one's *knowledge* changes then it is rational to change one's probability distribution accordingly. As we shall see, this answer requires dropping personalism, but we will see that is a price well worth paying.

If we turn now to Jeffrey conditionalization, we see that analogous questions arise, but answering them becomes that much more difficult. If 'evidence' propositions whose changes in probability are ones that rationally initiate updating, then which propositions are they? The answer that evidence propositions are the known propositions is not available. The new probabilities of these propositions are less than 1, whereas known propositions have probability equal to 1. Jeffrey conditionalization allows for the retention as epistemically possible (probability >0) propositions inconsistent with the evidence, but that makes no sense if the evidence

[8] See Williamson (2000:217) on the need for supplementing Bayesianism with an account of evidence.

propositions are known: propositions inconsistent with the subject's knowledge are not epistemically possible. That is really just to say that there is no natural place for an account of evidence as knowledge in Jeffrey conditionalization. However, that means that the explanation given above about why it is rational to update on changes in evidence ('because it is rational to update on changes in what one knows') is not available.

So Jeffrey conditionalization needs an account of evidence that allows evidence to be less than knowledge, yet explains why it is rational to update in response to changes in evidence. So far no such account is forthcoming. Jeffrey conditionalizers might aim to get by with something weaker: an account of evidence such that updating on evidence propositions is rationally permissible, even if not rationally required. Which propositions are those then? I suggest that something like the following thought is part of the Jeffrey conditionalizer's outlook. It would be rationally *impermissible* for a change in the probability of p to *initiate* updating, if that probability is fixed by its relationships to the probabilities of other propositions, via (F″) and (F‴) and the like. For a change in $P(p)$ in such a case would have to come about *as a result of* changes in those other probabilities. So the propositions that are permitted to initiate changes in a probability distribution are those whose probabilities are not dependent on the probabilities of other propositions. Einstein's GTR is a proposition whose probability depends on the probabilities of other propositions, concerning the outcomes of various experiments and observations as well as other theories, etc.; so it would not be permissible to update in response to a change in the probability accorded to GTR if that change comes about directly as opposed to as a result of changes in the probabilities accorded to those experiments, observations, and other theories. So it is not surprising that the propositions that Jeffrey and others use as examples for e in (J) are propositions arising from sensory perception: famously the subject who looks at a piece of cloth by candlelight and who is uncertain as to its colour and so gives various probabilities to the propositions that the cloth is green, that it is blue, and that it is violet (Jeffrey 1983:165).

Nonetheless, even those propositions do not meet the requirement that their probabilities are not dependent on the probabilities of other propositions. As Williamson (2000:217) points out, those probabilities will depend on what I may know about the reliability of my eyesight, the nature of the lighting conditions, the character of the cloth, etc. It is difficult to see that *any* proposition will meet this independence condition. One might think that reports of sensory appearances do meet the condition ('It seems to me that this cloth is probably green'). However, if these internal reports are fallible, then they will be dependent on my assessment of my own reliability in delivering such reports. On the other hand, if internal reports are infallible then we may ascribe probabilities of 1 to those reports, in which case we do not need Jeffrey conditionalization.

Jeffrey conditionalization looked liked a plausible option to take account of the fact that if personalism is true, then hardly ever, if at all, will a proposition have a probability of 1—certainly not the propositions of interest to science. Bayesian updating, whether strict or Jeffrey, needs an account of the circumstances in which a change in a probability distribution may rationally initiate updating of other propositions, since coherence can always be achieved by reversing the initial change. Such an account will come from outside the Bayesian framework. Strict conditionalization can be supplemented by a natural account of when that is: when a proposition

becomes known, that proposition may (and should) initiate updating.[9] No such account is available to the Jeffrey conditionalizer.

3.3 Knowledge and IBE

One important difference between Inference to the Best Explanation and standard Bayesianism is that the former tells us which propositions to infer from others—it includes a rule of detachment—whereas Bayesianism does not. So IBE as typically conceived might be summarised in a rule such as:

> (IBE) When A is a much lovelier potential explanation of the evidence than all its competitors (and it is in itself a good explanation, there is a good quantity of evidence, and alternative explanations have been sought assiduously), then infer that A is the actual explanation of the evidence and so is true.

Bayesianism might be supplemented by a rule of detachment—infer those propositions that when conditioned on the evidence reach a threshold probability (e.g. 0.95). But such rules are highly problematic unless the threshold is equal to 1.[10]

In this section I raise a related problem for compatibilism. Above I took it that the explanationist believes that using the rule (IBE) can lead to knowledge of the inferred proposition. This leads to the following problem in the Bayesian context if the following is true:

> In a typical explanationist inference, the evidence is consistent with more than one of the competing hypotheses, and no hypothesis has a prior equal to 1.

If this is correct, then a typical explanationist inference will produce a preferred hypothesis whose probability is less than 1. We saw above that (B′) entails that known propositions have a probability equal to 1. These together entail that a typical explanationist inference will not lead to knowledge. So explanationism seems to condemn us to scepticism. (Note that this problem is distinct from the challenge raised in 3.2. The latter arises from personalism. This problem is independent of personalism and arises simply from the probabilism in the form of (B′), even if we take the probabilities to be something other than credences, as this paper will propose.)

One might attempt to resolve this with the following. Bayes's theorem gives one the value of $P(h|e)$. Conditionalization says that on receipt of evidence e one should alter the probability attached to h to this value. However, if the epistemic circumstances are favourable then the subject gets to know, and the probability of the proposition rises to 1. That is, under such circumstances one can depart from conditionalization. This, however, is just a special case of the boost view of explanationism. While there are counterexamples to Bayesian conditionalization, this departure seems unmotivated.

Jeffrey Dunn (2014) raises the same problem in connection with evidence. Can a proposition that is learned by inductive inference itself be evidence? Let p be the proposition that a new Pope has been elected, and s be the proposition that

[9] For completeness, it should be added that updating should also be initiated when a proposition loses its status as knowledge.

[10] E.g. such a rule might permit the inference of each of two propositions individually but not their conjunction. In other cases, such as the proposition asserting that a given lottery ticket will not win, we might think that the proposition, though reaching the threshold, is not one that should be inferred.

there is white smoke coming from the Sistine Chapel. My credences are $P(p) = 0.5$, $P(s) = 0.5$, and $P(p|s) = 0.9$. I then see white smoke over the Sistine chapel, so s is now in my evidence set. So, with conditionalization, my credences are now $P(p) = 0.9$, $P(s) = 1$, and $P(p|s) = 0.9$. Can I now use p as evidence (e.g. with regard to some proposition concerning the immediate future of the Church)? According to Bayesianism, I cannot. For if p were now evidence, then it would have to have a probability of 1, whereas its probability is 0.9.

(Dunn concludes that p is not in my evidence, and so this makes trouble for E=K, since I can get to *know* that a new Pope has been elected by seeing white smoke. I don't think that this is really the problem. For as remarked above, the Bayesian is committed to E=K. If I really do know p, then p now becomes part of my background knowledge, and so, thanks to (B′), $P(p) = 1$. So the probability assigned by conditionalization, $P(p) = 0.9$ is inconsistent both with the proposition that p is now in my evidence and with the proposition that I now know p. I return to Dunn's argument below.)

3.4 Three challenges summarized

To summarize, the three challenges to compatibilism between Bayesianism and explanationism are:

1. Mere compatibilism with subjective Bayesianism answers Bayesian objections to IBE but fails to articulate how IBE can lead to the truth and so be a source of knowledge.

2. Bayesianism requires that evidence propositions and other known propositions have probability equal to 1. If that probability is a credence then, since no-one reasonably has credence 1 in most non-logical propositions, very few propositions will ever be known/evidence. The challenge is a consequence of Bayesianism commitment to personalism

3. If IBE selects among hypotheses that are all consistent with the evidence, Bayesian compatibilism implies that the inferred hypothesis has a probability of less than 1. But a known proposition cannot have a probability of less than 1. The challenge is a consequence of Bayesianism's commitment to probabilism plus the claim that IBE selects among hypotheses that are all consistent with the evidence.

I mooted that the explanationist might prefer some strong kind of objective Bayesianism. So in the next section I consider whether objective Bayesianism suffices to provide what we want for explanationist Bayesianism. I suggest that it does not. Objective Bayesianism will not suffice to answer 2, since objective Bayesians are personalists—the probabilities in question are credences. Below I consider a move beyond objective Bayesianism that rejects personalism and embraces a different conception of probability. This conception associates the probability of a hypothesis with its *plausibility*. I then return to challenge 3, which I answer by retaining probabilism but rejecting the assumption that IBE inevitably selects among hypotheses that are all consistent with the evidence.

4 Probability and plausibility

IBE requires objectivity if explanatory considerations are a guide to the truth. So we must rule out subjective Bayesianism. Should explanationism instead endorse objective Bayesianism, accepting further constraints on conditionalization? Standard approaches to objective Bayesianism will deploy versions of the principle of indifference, discussed above. Objective Bayesians will often hold that one's prior probabilities should be calibrated according to evidence of frequencies: if all one knows about some possible outcome o is that in the relevant circumstances such outcomes occur with frequency f, then one should accord o the probability f. Nonetheless, these kinds of constraint (indifference, calibration) do not suffice to capture what is needed to provide a form of objective Bayesianism that marries well with IBE. For often scientists are faced with competing hypotheses that are not neatly symmetric, for example: the hypothesis that the K-T extinction was caused by an asteroid collision versus the hypothesis that it was brought about by a period of rapid evolution. Furthermore, it is difficult to bring frequentist considerations to bear on such hypotheses—this is most especially obvious if the competing hypotheses concern the laws of nature. The other principles that objective Bayesians typically appeal to are, like indifference and calibration, principles that have an apriori appeal—Jaynes's (1968) principle of maximum entropy for example. The problem we faced was that compatibilist explanationism needs explanatory considerations to inform or parallel the use of Bayes's theorem in such a way that there is a correlation between higher posteriors and the truth. Further apriori principles seem too thin to be able to do that—if they did that would amount to an apriori justification of induction. So objective Bayesianism, as it is typically articulated, turns out to be an unsatisfactory answer to problem 1 after all.

We need therefore to move beyond objective Bayesianism's supplementation of standard Bayesianism with additional apriori principles. We may contemplate a 'super-objective' Bayesianism where good priors and likelihoods are determined by explanatory considerations that cannot be captured as apriori principles but instead have a *de facto* reliability. Super-objective Bayesianism would be a significant departure even from objective Bayesianism. If the relevant constraints on credences are apriori and assuming that credences, being doxastic attitudes, are introspectively knowable to the subject, then the subject is in a position to know apriori that her credences obey all the relevant constraints—the subject can know that she is being rational.[11]

So super-objectivism would undermine one motivation for personalism, that it permits the subject to assess the rationality of her evaluations. At the same time, super-objectivism does not itself solve problem 2, since that problem arises from the commitment to personalism. A natural step therefore is to adopt super-objectivism while renouncing personalism. Having renounced personalism, we need an alternative account of the evaluations of hypotheses. I agree with Salmon (1990:182) in holding that 'Plausibility considerations are pervasive in the sciences; they play a significant—indeed, *indispensable*—role.' Accordingly, I propose that our epistemic evaluations of hypotheses are evaluations of plausibility in the light of the evidence—such evaluations I shall call 'plausibilities'. Salmon's objective Bayesianism is an important attempt to link Bayesian confirmation to Kuhn's five criteria of

[11] I would deny that a subject is always able to know her own credences, but the thought that credences are transparent to introspection and that Bayesianism allows us to police our own rationality is a common one (see Savage 1954:20; Egan and Elga 2005).

theory choice via plausibility considerations, to which the proposals in this paper have some similarity. Salmon, however, regards plausibility arguments as supplying priors within the context of personalism, whereas I follow Williamson (2000) in regarding plausibilities as alternative to credences. (I also regard plausibility considerations as determining the likelihood and the expectedness—or, directly, their ratio.)

We retain probablism regarding plausibilities: the plausibility of a hypothesis in the light of the evidence can indeed be modelled by probabilities.[12] Insofar as we can quantify plausibilities, we can represent them as ranging from a minimum, 0, the plausibility of a hypothesis inconsistent with the evidence, to a maximum, 1, the plausibility of a hypothesis entailed by the evidence. Thus represented the plausibility of a disjunction of mutually exclusive hypotheses is the sum of the plausibilities of the hypotheses individually; in particular, the plausibility of a disjunction of a hypothesis and its negation is maximal. Thus finite additivity is satisfied.[13] If my evidence allows me to know, for example, that the Earth orbits the Sun, then that hypothesis has maximum plausibility and could not gain additional plausibility from additional evidence. Likewise, if my theory entails that water is frozen at 80°, then my evidence shows that this theory has zero plausibility. Since plausibilities can be given values of 0 and 1, this approach avoids problem 2.

Above I identified two aspects to the explanatory loveliness of a hypothesis: an internal component, which is independent of the evidence, and an external component, which concerns the ability of the hypothesis to explain the evidence in question. The internal component corresponded to the prior probability of the hypothesis, that is to say, the plausibility of the hypothesis independent of its evidential support. (The 'internal/external' terminology here does *not* align with a distinction between factors that are acceptable to an internalist epistemology and those acceptable only to an externalist epistemology.[14]) In science we do find assessments of plausibility of this sort. For example, scientists in the eighteenth century wondered about the force of attraction and repulsion between charged objects. In 1785, Charles-Augustin de Coulomb proposed an inverse square law. Even before the torsion balance experiments that he then carried out, this hypothesis carried a high degree of (evidence-independent) plausibility because of its analogy to the inverse square law of gravitation. This is evidence-independent plausibility in that it is not the product of any process of conditionalization (e.g. on the law of gravitation), not even in an idealised way. Why should we not treat the structural similarity between Coulomb's hypothesis and Newton's law as a piece of evidence (since that similarity is a known fact) in favour of the former? We don't typically think in that way—in the context of the heuristic approach this fact is psychologically most plausibly seen as influencing our prior for Coulomb's hypothesis. It would be difficult to implement Bayesianism with this as evidence. For that would require setting an independent

[12] This approach therefore builds on Williamson's (2000) notion of *evidential probability*.

[13] A further question asks how plausibilities relate to credences. Clearly there is no neat relationship, in that some propositions will have a plausibility of 1 but a credence of less than 1. On the other hand, if a subjective Bayesian does treat a subject's evidence/knowledge as having credence equal to 1 and has priors coinciding with those of the plausibility theorist, then conditionalization will lead to the same posterior probabilities.

[14] This internal explanatory virtue is what Williamson (2000:211) calls the 'intrinsic plausibility [of the hypothesis] prior to investigation'. It does not seem correct, though, that plausibility can be a truly *intrinsic* property of a hypothesis. If it were, there would be a plausibility that is independent of the psychological make-up of even an ideal human scientist, including their innate psychology. Such a notion looks as if it would be close to the apriori objectivism we considered but rejected above.

prior for the hypothesis, and where does that come from? And it would require setting a prior for the evidence, the similarity between Coulomb's hypothesis and Newton's law, and it is also unclear where that could come from. This influence on the plausibility of Coulomb's proposed law can come about even if we hold that the law of gravitation is not itself known (and so not evidence). More generally, experience can influence the assessment of plausibility in ways that cannot simply be reduced to an evidential relationship (rather as one may correctly recognize the facial similarity of two members of the same family without being able to articulate the basis of that similarity).So a hypothesis can gain plausibility by its connection to elements of the scientific context in a way that is independent of conditionalization on evidence. Analogy with well-confirmed hypotheses is an example of how this occurs. And so we have one answer to the question: where does evidence-independent plausibility come from? (There may be other sources of evidence-independent plausibility.[15]) The remaining question concerns the objectivity of this plausibility.

5 Assessing objective plausibility

I have just argued that we make judgments of plausibility of a hypothesis that are independent of its conditionalization on the evidence—in so doing we respond to internal explanatory virtues, and these may be influenced, for example, by analogy with other theories. Similar remarks may be made about external explanatory virtues, which allow us to assess the factor by which the degree of plausibility changes when the evidence is obtained (i.e. the ratio $P(e|h)/P(e)$). These two may be informed by exemplars of successful explanatory relations between theories and evidence.

Indeed, we can draw on a central lesson of the work of Thomas Kuhn (1970), which is that scientists' assessment of any hypothesis is crucially influenced by their prior experience of other successful theories.[16] According to Kuhn, such theories, which he called 'exemplars' or 'paradigms', are models that inform the development and assessment of newer hypotheses. Because Newton's theory of gravitation was so successful, it became a model (in multiple ways) of what a good theory might look like, and so, as we saw, Coulomb's proposed law acquired plausibility because of its similarity to that model.

We can link this point to explanationism with the following proposal from Walker (2012). Scientists acquire their sense of explanatory virtues from their training with and use of exemplars. Let us assume some degree of scientific realism—the theories in question have a high degree of truth-content. Let us also assume that theories similar to true or highly truthlike theories are more likely to be truthlike than less similar theories. If these assumptions are correct (and they are themselves at least plausible), then the practice of allowing one's preferences among theories to be influenced by their explanatory virtues will be a truth-conducive practice. Exemplars change over time, and so the explanatory virtues acknowledged by a community will

[15]For example, another source might be the epistemological conservatism advocated by Poston (2014). That a proposition is believed by a subject suffices to give it some degree of plausibility. Poston provides arguments for conservatism from an internalist perspective, while I can see externalist, naturalistic reasons in its favour.

[16]We can accept this result of Kuhn's work without taking it to have any anti-realist consequences, as this section argues. See Bird (2000). Salmon (1990:196–9) does likewise, though concentrates on Kuhn's five shared criteria of theory choice, which Salmon takes to play a role in determining the plausibilities of hypotheses and thereby the priors of those hypotheses.

change over time also—we can see this as a process of the evolution of a scientific community's values in the direction of greater reliability.[17] We should not expect these values and their applications to be the same in all scientific fields. Different fields have different exemplars and so values (or, more frequently, what counts as instantiating a given value) may differ between fields.[18]

So now let us imagine a scientist who considers some candidate hypothesis in advance (or independently) of the evidence. Her assessment of its plausibility is influenced in part by the degree to which it possesses the explanatory virtues acknowledged by the community. In the light of the forgoing, and if the two assumptions are correct, we may expect there to be some correlation between the plausibility assessments and truth. Consequently, probabilities interpreted as (quantified) plausibilities will be objective in the sense relevant to challenge 1: probabilities will relate to truth, in that hypotheses assessed as more plausible will be more likely to be true.

Note that it does not need to be claimed that this correlation is perfect. From the Bayesian perspective the assessment of prior plausibility might be only rough and ready and still good enough. For the phenomenon of the washing-out of priors will mean that imperfections in the prior probability will be increasingly insignificant as ever more evidence comes in. The washing-out of priors argument does not help the subjective Bayesian since that position accepts no epistemologically relevant constraints on priors, meaning that priors are acceptable that are so extreme that they cannot be washed out by a reasonable quantity of evidence. If plausibility is a rational constraint on priors, then extreme priors will be excluded (unless warranted in particular cases by extreme plausibility or implausibility).

6 Plausibility, knowledge, and eliminativism

Since plausibilities are not credences, a subject who assigns evidential probability of 1 to a proposition is not obliged to accept wild bets on the truth of that proposition. So the second challenge is not a problem for this approach. A probability of 1 means that the proposition is maximally plausible *given the evidence and background knowledge*. Rationally assessing a proposition as maximally plausible does not seem unachievable in the way that a rational credence of 1 looks unachievable for most propositions.

However, $P(h) = 1$ does mean that any proposition inconsistent with h (e.g. competitor hypotheses) must have zero probability, that is to say maximum implausibility in the light of the evidence. This looks to be in tension with IBE, since the latter is typically understood as making inferences on the basis of one explanation being a *better* explanation than the competitor hypotheses. The latter implies that one can make a reasonable IBE when the best explanation is just rather *more* plausible than its competitors; the latter thus retain some degree of plausibility. If that is right, then the inferred proposition will not be knowledge, since knowing h requires $P(h) = 1$ and $P(h') = 0$ for all competitor hypotheses h'.[19]

[17] By 'values' here I mean the features of a theory that the community regard as virtues.

[18] As Salmon (1990:186) remarks, 'simplicity ... [is] a highly relevant characteristic, but one whose applicability varies from one scientific context to another'.

[19] Since a subject's known propositions and evidence propositions are the same, and evidence propositions have a plausibility of 1, so also do known propositions. $P(h) = 1$ is consistent with fallibilism, when the probabilities are plausibilities. If the probabilities are *credences* then fallibilists will want probabilities to be less than 1 even for knowledge.

So knowledge of a proposition inferred on the basis of explanationist considerations requires that the competing hypotheses must be *ruled out* by the evidence and background knowledge: what may be called *eliminative abduction*. It might be wondered whether explanationist considerations can ever achieve this. Perhaps the most that can be achieved is that a hypothesis is shown to be highly implausible.

This challenge cannot be covered comprehensively here. But I can indicate some of the lines of response I advocate:

- Scientists do think that hypotheses can be shown to be false: Harvey showed that blood could not continuously be produced afresh by the liver; Galileo's observations of the phases of Venus proved that the Sun and Venus did not both orbit the Earth; Pasteur's experiments with swan-necked flasks refuted the hypothesis spontaneous generation was the cause of the spoiling of organic liquids such as broth or wine[20]; Rutherford's gold-foil experiments disproved the plum-pudding model of the atom; Wu's experiment ruled out the conservation of parity; and so forth. The pre-theoretic judgment of scientists is that in cases such as these it is possible to know that a hypothesis is false.

- That pre-theoretic judgment is challenged by the use of Duhem's contention that use of an experiment to refute a hypothesis will typically depend on the use of auxiliary hypotheses. All one can know, it is said, is that the conjunction of the hypothesis under test *and* the various auxiliary hypotheses is false, not that that the test hypothesis itself is false. While such an argument appears to *lead to* scepticism, it is instead the *product* of an antecedent scepticism. For the conclusion holds only if the auxiliary hypotheses are themselves not known. If a scientist does know the auxiliary hypotheses to be true, then she can know that the test hypothesis is false. So a sceptical conclusion about the possibility of refutation can only be drawn if one is already sceptical about the possibility of knowing the truth of auxiliary hypotheses. While this may be problematic for Popper, since he is a sceptic of that sort, there is no such problem for those who are not already sceptical about the possibility of scientific knowledge. The truth of an auxiliary hypothesis may always be disputed. But that fact does not show that the hypothesis is not known by any scientist—immunity from challenge is not a condition on knowledge.

- Even if one is not antecedently a sceptic, one might think that sceptical hypotheses are among the competitor hypotheses that would need to be refuted in order for the favoured hypothesis to have a probability of 1, and that the very nature of such hypotheses is that they cannot be refuted. For example, if one is considering competing explanations of the evidence *e*, is not one of the hypotheses to be considered the hypothesis that an evil demon has made it falsely appear that *e* is the case? If so, how would one refute that hypothesis? I say that *that* hypothesis need not be considered. For it is inconsistent with and so ruled out by *e*. An explanation of why things falsely appear is if *e* is necessarily not an explanation of the evidence *e*. *Ex hypothesi* our evidence is *e* not merely the proposition that things appear as if *e*. If one responds that the former should not be included in our evidence, and only the latter propo-

[20] Pasteur is reported (Vallery-Radot 1926:109) to have concluded a public lecture at the Sorbonne in 1864 by saying "Never will the doctrine of spontaneous generation recover from the mortal blow of this simple experiment."

sition, then that is to adopt an antecedent scepticism that the sort of propositions we take to be our evidence are never known.[21]

- One might nonetheless remain concerned that there are yet other sceptical hypotheses consistent with our total evidence (as non-sceptically conceived). For example, we might propose the following hypothesis for the outcome of Pasteur's swan-necked flask experiments: there is spontaneous generation in the flasks but the evil demon uses special powers to kill off the organisms thus generated before they can spoil the organic infusion. So Pasteur's evidence is just as we take it to be (there is no perceptual illusion caused by the evil demon). To this kind the sceptical hypothesis, the following responses are worth considering:
  - It is not entirely clear that such hypotheses really are consistent with our evidence. The existence of an evil demon with a power of unspecified mechanism is hypothesized. If some specific mechanism were hypothesized, then it might be clear that what we know from physics and chemistry and such like is inconsistent with the hypothesis. Failure to specify the mechanism doesn't remove concerns about inconsistency, for it is far from obvious that there is some possible mechanism consistent with our scientific evidence.
  - Even if such possibilities are consistent with our evidence, we may still be entitled either to ignore them or to give them zero plausibility. For such possibilities are remote possibilities, in the sense that they are true only in remote possible worlds.[22] And it may be sufficient for eliminative knowledge of the kind being proposed here that all competing hypotheses true in close worlds are eliminated by the evidence. This might be endorsed by those who take knowledge to be tantamount to satisfying a requirement of safety (roughly: S knows that p iff S could not easily be mistaken in believing that p). Some contextualists will endorse a similar proposal, that in the scientific context it is proper to ignore (not consider) such hypotheses. (Since the probabilities are not credences but plausibilities, a contingent but remote possibility can have a probability equal to zero.)

In addition to being able to rule out hypotheses, the subject needs also to know that the true hypothesis is among the competing hypotheses she considers. While a subject can be wrong about this (Stanford 2006), I have argued elsewhere that such knowledge is certainly possible (Bird 2005): in a popular example, the detective can know that (at least) one of the fifteen passengers on the train committed the crime, but examples also abound in both daily life and in science. Returning to Dunn's case,

[21] Of course, the sceptic may ask of some evidence proposition e how we can be sure that e is in our evidence—perhaps it only appears to be evidence? Can we distinguish between genuine and merely apparent evidence? While we may be able in many cases to make that distinction, the important point is that for present purposes it does not need to be answered. Being able to demonstrate that e is in S's evidence is not a condition of its in fact being in S's evidence—the KK principle is rejected. Our current question asks which hypotheses can be ruled out by a subject's evidence, not which hypotheses can be ruled out by what a subject knows to be her evidence.

[22] The condition that the possibilities in question are remote, not close is important. One cannot ignore an implausible hypothesis that is true in a close world. Police may suspect a known burglar who is suddenly rich of having pulled off a heist. But they should not ignore the possibility of a lottery win, since such a possibility (if the burglar bought a ticket) is a close possibility even though it is an improbable one.

if the subject knows that a new Pope has been elected when she sees white smoke above the Sistine Chapel, then it must also be the case that the subject knows that if there is white smoke above the Sistine Chapel then a new Pope has been elected. (This could well be a case of eliminative abduction: another explanation for the smoke is that there is an accidental fire. But if so, the smoke would not be white and there would be fire alarms. So that alternative hypothesis is ruled out by the evidence.) If the subject does know that conditional proposition, then her conditional probability $P(p|s) = 1$, not 0.9 as in Dunn's example. In which case, $P(p) = 1$, and Dunn's problem disappears; there is no Bayesian problem with p being evidence. On the other hand, if $P(p|s) = 0.9$ then the subject does not know that if there is white smoke then there is a new Pope, in which case acquiring the evidence that there is white smoke cannot lead to knowledge that there is a new Pope (and so that proposition is not an evidence proposition). Either way, E=K is safe. More importantly, in the first case, where $P(p|s) = 1$ and so $P(p) = 1$, the inferred proposition is known (and so is evidence and has probability = 1).

7 Conclusion

The heuristic approach to compatibilism between Bayesianism and IBE can see off the challenge from the subjective Bayesian. Explanationist considerations do not supplement Bayesian considerations by giving a boost to the probability of a hypothesis when conditionalized on the evidence. Rather we can see explanationist factors as corresponding to the various elements of Bayes's theorem, in particular the prior $P(h)$ and the ratio of the likelihood to the expectation of the evidence, $P(e|h)/P(e)$. Still, that cannot be all that explanationist considerations achieve, for IBE reasonably aspires to give us knowledge, whereas subjective Bayesianism does not. This may motivate a move from subjective to objective Bayesianism. But that position faces the problem that no scientific proposition can reasonably be accorded credence of 1, in which case no proposition can ever be known or be evidence. So we need to move beyond the personalism of both subjective and objective Bayesianism: We should not understand the probabilities with which we are concerned as credences but rather as something else. That something else must allow it to be possible to have evidence and knowledge and to accord the relevant propositions probability 1. Williamson's evidential probability does this. But that requires giving an account of what that probability is: what feature of a subjects epistemic state does evidential probability aim to capture? The proposal is that it captures assessments of plausibility in the light of the evidence.

Consideration of how a hypothesis acquires plausibility on explanationist grounds (e.g. explanationist values that feed into plausibility are generated by assessments of similarity to successful, highly truthlike theories) shows that it can be objective (in the relevant sense) and may be correlate with the truth. IBE does face an additional problem in that if it is to lead to knowledge, explanationist considerations must in fact be able to rule out competing hypotheses i.e. reduce their plausibility to 0. While showing that takes us beyond this paper, I have argued that so doing does not face any obvious and insurmountable obstacle.[23]

[23] My thanks to Kevin McCain, Richard Pettigrew, Ted Poston, and Timothy Williamson for their instructive discussion, comments, and advice.

References

Bird, A. 2000. *Thomas Kuhn.* Chesham: Acumen.

Bird, A. 2005. Abductive knowledge and Holmesian inference. In T. S. Gendler and J. Hawthorne (Eds.), *Oxford Studies in Epistemology*, pp. 1–31. Oxford: Oxford University Press.

Briggs, R. 2009. Distorted reflection. *Philosophical Review* **118**: 59–85.

Dunn, J. 2014. Inferential evidence. *American Philosophical Quarterly* **51**: 201–13.

Egan, A. and A. Elga 2005. I can't belive I'm stupid. *Philosophical Perspectives* **19**: 77–93.

Einstein, A. 1991. *Autobiographical Notes.* La Salle IL: Open Court.

Horwich, P. 1982. *Probability and Evidence.* Cambridge: Cambridge University Press.

Huemer, M. 2009. Explanationist aid for the theory of inductive logic. *British Journal for the Philosophy of Science* **60**: 345–75.

Jaynes, E. 1968. Prior probabilities. *Institute of Electrical and Electronic Engineers Transactions on Systems Science and Cybernetics* **SSC-4**: 227–41.

Jeffrey, R. 1983. *The Logic of Decision* (2nd ed.). Chicago: University of Chicago Press.

Joyce, J. M. 1998. A nonpragmatic vindication of probabilism. *Philosophy of Science* **65**: 575–603.

Keynes, J. M. 1921. *A Treatise on Probability.* London: Macmillan.

Kuhn, T. S. 1970. *The Structure of Scientific Revolutions* (2nd ed.). Chicago, IL: University of Chicago Press.

Leitgeb, H. and R. Pettigrew 2010a. An objective justification of Bayesianism I: Measuring inaccuracy. *Philosophy of Science* **77**: 201–35.

Leitgeb, H. and R. Pettigrew 2010b. An objective justification of Bayesianism II: The consequences of minimizing inaccuracy. *Philosophy of Science* **77**: 236—72.

Lipton, P. 2004. *Inference to the Best Explanation* (2nd ed.). London: Routledge.

Mahtani, A. 2012. Diachronic Dutch book arguments. *Philosophical Review* **121**: 443–50.

Milne, P. 1996. log [P(h/eb)/P(h/b)] is the one true measure of confirmation. *Philosophy of Science* **63**: 21–6.

Poston, T. 2014. *Reason and Explanation.* Basingstoke: Palgrave Macmillan.

Psillos, S. 2004. Inference to the best explanation and Bayesianism. In F. Stadler (Ed.), *Induction and Deduction in the Sciences*, pp. 83–91. Kluwer.

Salmon, W. C. 1990. Rationality and objectivity in science or Tom Kuhn meets Tom Bayes. In C. W. Savage (Ed.), *Scientific Theories*, Volume 14 of *Minnesota Studies in the Philosophy of Science*, pp. 175–204. Minneapolis, MN: University of Minnesota Press.

Salmon, W. C. 2001. Explanation and confirmation: A Bayesian critique of inference to the best explanation. In G. Hon and S. S. Rakover (Eds.), *Explanation: Theoretical Approaches and Applications*, pp. 61–91. Dordrecht: Kluwer.

Savage, L. J. 1954. *The Foundations of Statistics*. New York NY: Dover.

Schupbach, J. N. 2016. Inference to the best explanation, cleaned up and made respectable. In K. McCain and T. Poston (Eds.), *Best Explanations: New Essays on Inference to the Best Explanation*. Oxford: Oxford University Press. (this volume).

Stanford, P. K. 2006. *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. New York: Oxford University Press.

Vallery-Radot, R. 1926. *The Life of Pasteur*. Garden City, NY: Garden City Publishing Company. Translated by Mrs R. L. Devonshire.

van Fraassen, B. 1989. *Laws and Symmetry*. Oxford: Oxford University Press.

Walker, D. 2012. A Kuhnian defence of inference to the best explanation. *Studies in History and Philosophy of Science* **43**: 64–73.

Williamson, J. 2010. *In Defence of Objective Bayesianism*. Oxford: Oxford University Press.

Williamson, T. 1997. Knowledge as evidence. *Mind* **106**: 1–25.

Williamson, T. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.