Locating IBE in the Bayesian Framework*

Jonathan Weisberg
University of Toronto

Abstract Inference to the Best Explanation (IBE) and Bayesianism are our two most prominent theories of scientific inference. Are they compatible? Van Fraassen famously argued that they are not, concluding that IBE must be wrong since Bayesianism is right. Writers since then, from both the Bayesian and explanationist camps, have usually considered van Fraassen's argument to be misguided, and have plumped for the view that Bayesianism and IBE are actually compatible. I argue that van Fraassen's argument is actually not so misguided, and that it causes more trouble for compatibilists than is typically thought. Bayesianism in its dominant, subjectivist form, can only be made compatible with IBE if IBE is made subservient to conditionalization in a way that robs IBE of much of its substance and interest. If Bayesianism and IBE are to be fit together, I argue, a strongly objective Bayesianism is the preferred option. I go on to sketch this objectivist, IBE-based Bayesianism, and offer some preliminary suggestions for its development.

1 Introduction

Recent literature on Inference to the Best Explanation (IBE) has focused on whether and how IBE fits with the Bayesian view of non-deductive reasoning. Van Fraassen (1989) famously argued that IBE is incompatible with Bayesianism because IBE requires us to violate the Bayesian rule of conditionalization. Van Fraassen concludes that IBE is bad epistemology because violating conditionalization is irrational. Responding to this argument, explanationists have mostly questioned the assumption that IBE really does clash with conditionalization. Thus a sort of compatibilism has emerged as the way to go.[1] In the words of Peter Lipton, "My objection to the argument that Inference to the Best Explanation is wrong because Bayesianism is right [is that] the argument is a non-sequitur, because Bayesianism and Inference to the Best Explanation are broadly compatible." (Lipton 2004: p. 106) Lipton

* I am extremely grateful to Frank Arntzenius, Alexander Bird, Anjan Chakravartty, Barry Loewer, Chris Meacham, Samir Okasha, Adam Sennet, and two anonymous referees for *Synthese* for much helpful discussion. I am also indebted to the attendees of the 2007 Annual Philosophy of Science Conference in Dubrovnik, Croatia, and the 2007 Annual Conference of the British Society for the Philosophy of Science in Bristol, UK, for their helpful comments and suggestions.

1 Not without exception, e.g. Douven (1999) and Tregear (2004).

goes on to suggest that the two views may even be complementary, and that the Bayesian and the explanationist should be friends.

Unfortunately for compatibilists, however, IBE cannot be made consistent with conditionalization, at least not without sapping much of the substance from IBE. To a first approximation, the problem is this: IBE tracks explanatory goodness, while conditionalization tracks prior conditional degree of belief. For the two rules to agree, 'explanatory goodness' must be understood in a way that makes it subservient to prior conditional degree of belief, which robs IBE of much of its appeal and interest.

While my first aim here is to revive and elaborate van Fraassen's argument that IBE and conditionalization are incompatible, this is not because I have something against IBE. Rather, reviving the incompatibility argument is the first step in an attempt to push compatibilists towards collaboration with the objective Bayesian rather than the subjective Bayesian. Such a collaboration strikes me as a more promising avenue for both explanationists and Bayesians. Forced to choose between IBE and subjective Bayesianism, I hope that compatibilists will reject subjectivism and pursue a Bayesian IBE with a more objectivist flavor. The next step is then to make compatibility with objective Bayesianism attractive, both to the explanationist and to the objective Bayesian. To that end, I close with some preliminary suggestions for developing explanationism in the objective Bayesian framework.

2 The Incompatibility Argument

In its most naive form, IBE tells us to believe the best explanation as the true one. But this formulation is too naive, neglecting many important choice-points for IBE's proponents. What makes an explanation better? Should we accept the explanation that actually is best, or just the one that we think is best? What if there may be explanations we haven't thought of yet? What if the best explanation isn't very good?

In the interest of simplicity and generality, I will be suppressing these complications as much as possible. What makes for explanatory goodness will not matter so long as we may assume that our account of goodness must agree with clear-cut, intuitive verdicts about particular cases (but see the third objection to the burglar case on p. 8). The question whether we should favor the best explanation, as opposed to the one we think best, will intrude on the discussion briefly, but will be seen not to affect the argument (p. 8). The issue of as-yet unconceived explanations will make an oblique appearance when we discuss cases of non-existent priors (p. 10), but is otherwise orthogonal to our concerns here. And the "what if the best isn't very good?" concern no longer applies once IBE is translated into

degree-of-belief terms, which we are about to do.

As stated, IBE is a claim about what we ought to believe, but bare bones Bayesianism has nothing to say about qualitative belief; it deals solely in degrees of belief. To compare the two views we have to either introduce qualitative belief into the Bayesian framework, or recast IBE in terms of degrees of belief. Some Bayesians reject the notion of qualitative belief outright (Jeffrey 1970; Christensen 2004), so a natural place to start is by casting IBE in terms of degrees of belief. This is certainly not the only place we might start. Another natural starting point is to use explanatory reasoning as a factor that, together with degrees of belief, determines qualitative belief, a la (Levi 1980) and (Maher 1993). But van Fraassen and his compatibilist respondents have, by and large, preferred to see IBE as influencing degrees of belief, so we will pursue this avenue.

Van Fraassen (1989) interprets IBE as a policy of favoring explanatorily superior hypotheses by giving their subjective probabilities a post-conditionalization boost. To see how that works, we first need to appreciate van Fraassen's brand of Bayesianism, and the understanding of conditionalization that comes along with it.

On the subjective Bayesian view, a rational agent starts with degrees of belief represented by a probability function, p. When she receives new evidence E, the subjective Bayesian requires her to follow this rule:

Subjectivist Conditionalization When you gain new evidence E, your new degree of belief in a hypothesis H, call it $q(H)$, should be your old degree of belief in H conditional on E: $q(H) = p(H|E)$.

Subjectivist Conditionalization makes your new degree of belief in a hypothesis a function of two things: the new evidence and your conditional degree of belief prior to acquiring that evidence. This is a "subjective" policy in two senses. First, probabilities here are understood as the subjective state of an agent, namely her degrees of belief. Second, to the extent that rational agents may have different prior degrees of belief, there is room for rational inter-subjective disagreement. If you rationally hold degree of belief $p(H|E) = x$, while I rationally hold $p'(H|E) \neq x$, then we will be rationally obliged to draw different conclusions given evidence E.

Bayesians who take the foregoing description of epistemic rationality to be fairly complete are of the more radically subjective kind. They require only that rational agents have probabilistic degrees of belief and update according to Subjectivist Conditionalization, which leaves plenty of room for inter-subjective disagreement. More conservative subjective Bayesians will add requirements to the list. Van Fraassen, for example, requires that rational agents obey a principle of

Reflection (van Fraassen 1984, 1995),[2] and many others impose a chance-credence constraint like Lewis's Principal Principle (Lewis 1980). As further constraints are added, the range of rationally permissible starting points is narrowed, and room for inter-subjective disagreement is eliminated. More constraints thus means a more objective brand of Bayesianism.

The kind of Bayesian who concerns us for now lives towards the subjective end of this subjective-objective continuum. She only requires that rational agents obey the probability axioms, maybe something along the lines of the Reflection and Principal Principles, and perhaps a few other fairly weak constraints; beyond those synchronic constraints, Subjectivist Conditionalization is the only thing she requires. This is the kind of Bayesian that van Fraassen is, but more importantly, it is the kind of Bayesian to whom his incompatibility argument will matter. The crux of the argument is going to be that IBE requires violations of Subjectivist Conditionalization, and this result will not bother the strongly objective Bayesian, since she will require violations of Subjectivist Conditionalization as a matter of course. Anyone whose prior degrees of belief violate her strongly objective constraints will be expected to alter her degrees of belief, in violation of Subjectivist Conditionalization, to bring them into line. So the more objective Bayesian will not be troubled if IBE requires violations of Subjectivist Conditionalization, since she already sees violating Subjectivist Conditionalization as legitimate when necessary in order to respect other principles she holds.[3] Ultimately I will be encouraging explanationists to pursue this more objective brand of Bayesianism. But for now let's focus on the more subjective Bayesian, to see to what extent she can be friends with the explanationist.

Returning to van Fraassen's argument, what does IBE amount to in Bayesian terms? As van Fraassen renders it, IBE amounts to applying Subjectivist Conditionalization in response to new evidence E, and then boosting the probability of those hypotheses that better explain E, while penalizing those that provide a poorer explanation. The net result is that $q(H)$ will not be identical to $p(H|E)$, but will instead be a variation on $p(H|E)$ that gives more or less weight to H according as it is a better or worse explanation of E. Van Fraassen then points out that someone who applies IBE in this way is open to a Lewis-Teller style Dutch

2 In fact, van Fraassen (1995; 1999) takes Reflection to be a more fundamental requirement than Subjectivist Conditionalization, and regards Subjectivist Conditionalization as only required under certain (usually satisfied) conditions. I will be ignoring this wrinkle in van Fraassen's view for the purposes of simplicity.

3 There is a possible kind of objective Bayesian who will prefer that you violate her other constraints than violate Subjectivist Conditionalization. I doubt that many actual objectivists hold this view, but if they do, then they are, for my purposes, subjectivists. To attain compatibility with this kind of Bayesian, the explanationist will have to make concessions like those I outline for the explanationist who seeks compatibility with subjective Bayesianism.

book, as is anyone who violates Subjectivist Conditionalization (Teller 1973). Since Dutch-bookability is supposed to be a sign of irrationality, van Fraassen concludes that IBE is an irrational policy. Diachronic Dutch book arguments are not as well-regarded as they once were, but the really important thing is that IBE, as van Fraassen renders it, leads to violations of Subjectivist Conditionalization. So any argument that Subjectivist Conditionalization is rationally required is an argument against IBE.[4]

The explanationist has two options here. She can either accept that IBE violates Subjectivist Conditionalization, or she can deny that there really is a conflict. The latter move has been most popular, with apologists arguing that van Fraassen has misconstrued IBE, and that the inferences prescribed by IBE are really the same as those required by Subjectivist Conditionalization. IBE only requires that better explanations end up with higher posterior probabilities, and this might happen *as a result of applying Subjectivist Conditionalization*, rather than in violation of Subjectivist Conditionalization. Van Fraassen supposes that the preference for better explanations must manifest itself in an extra, non-Bayesian step after Subjectivist Conditionalization is applied. But why couldn't the very application of Subjectivist Conditionalization be what ensures higher posterior credence in the better explanation?

To illustrate, let H be the hypothesis that the coin I hold is evenly balanced, H' the hypothesis that it is weighted towards heads, and E the fact that five of the next ten tosses come up heads. Presumably, H is the better explanation of E, and so it ought to end up with a higher degree of belief. Assuming that both hypotheses are equally plausible to begin with, and that H confers a higher probability on E than H' does, as seems reasonable, then the prior conditional probabilities will be such that $p(H|E) > p(H'|E)$. Thus applying Subjectivist Conditionalization will result in $q(H) > q(H')$, just as IBE says should happen. Far from requiring violations of Subjectivist Conditionalization, says the compatibilist, IBE is actually enforced by Subjectivist Conditionalization.

3 The Limits of Compatibilism

As a response to van Fraassen's challenge, compatibilism depends crucially on a tight correlation between an explanation's superiority and its prior conditional degree of belief. In the coin-flip example, applying Subjectivist Conditionalization yielded a higher posterior degree of belief for H because H started with a higher

4 Other arguments for Subjectivist Conditionalization have been made by appeal to "rigidity" (Jeffrey 1983), cognitive decision theory (Greaves & Wallace 2006), and the Reflection Principle (van Fraassen 1999). Lange (1999) offers a calibration-based argument for a form of conditionalization, but not for Subjectivist Conditionalization.

conditional degree of belief. If instead we had had $p(H|E) < p(H'|E)$, applying Subjectivist Conditionalization would have yielded the explanationist-unfriendly result that $q(H) < q(H')$. That is, after all, the content of Subjectivist Conditionalization. The agreement between prior conditional credence and explanatory superiority is thus crucial; to the extent that these two factors diverge, so will IBE and Subjectivist Conditionalization.

The problem I want to raise for compatibilists is that prior conditional credence and explanatory goodness do diverge. I will begin by defending the modest claim that it is *possible* for them to diverge, and address some preliminary objections. I will then address what I take to be the predominant explanationist view: that IBE is a heuristic we use to approximate the ideal that Bayesian standards represent, so possible divergences can be ignored so long as they are the exception and not the rule. I will argue that this heuristic take on IBE makes it inapplicable in important cases, most notably in cases that avowed compatibilists cite as exemplars of IBE. It also forces IBE to take on the radical subjectivity of subjective Bayesianism, diminishing one of IBE's central appeals, namely its agreement with common sense and paradigm instances of good scientific inference.

3.1 Possible Divergence

To see how it is *possible* for Subjectivist Conditionalization and IBE to diverge, consider a case where IBE yields a clear-cut verdict. Suppose you come home one day to find the front door open and the lock broken. Furniture is overturned, the contents of the shelves are on the floor, and valuables are missing. One explanation is that someone broke in and stole your belongings, making a mess in the hurried process. But here is a second possible explanation. One burglar broke the lock and entered your house, only to encounter another burglar, who had found his way in through a window just a few minutes earlier. The two fought, making a mess in the process, before a police officer entered, having noticed the broken lock from the street. The two burglars took off, and the police officer, deciding to take advantage of the situation rather than risk having it revealed that he failed to apprehend either burglar, stole your belongings.

Whatever your account of explanatory virtue, if IBE applies here, it surely favors the first explanation.[5] We might say that this is because it is simpler, or more unifying, or something else, but we can leave that open. However we spell out its virtues, the important thing is that the first explanation would still be the more virtuous one even if your prior conditional degree of belief were higher for the second explanation. What makes the second explanation unvirtuous is

5 If you think IBE does not apply here, substitute a case where it does; the same critical discussion will apply.

the intricacy of the narrative. It is the multiplicity of characters, entries, exists, and interactions that makes it a ragged explanation, and these features are present however we try to fit them under rubrics like 'simplicity' and 'unification'. So explanatory goodness, whatever it is, looks to be at least somewhat independent of prior conditional degree of belief.

You might well think that this is too quick, and that the subtleties of IBE are not being appreciated here. I will try to address these concerns by raising and responding to three objections.

First Objection: explanatory goodness is relative to background belief. An explanation's goodness must be evaluated in the presence of background assumptions, so can't we say that explanatory goodness actually does depend on prior degree of belief, contrary to appearances? To be sure, we can imagine background assumptions that would make the second explanation the better one. If burglars target the same houses so often that run-ins are more frequent than not, and if police corruption is rampant in your city, then the second explanation might be better. But notice that, even if explanatory virtue depends on your background beliefs, it does not depend on the particular belief-state relevant to Subjectivist Conditionalization, namely your *prior conditional credence* in the explanation given the evidence. It is consistent with everything that the subjective Bayesian requires of you to fully believe that police corruption is extremely rare, that burglars almost never run into one another on the job, etc., and yet have a much higher conditional degree of belief in the two-burglar explanation. Nothing about the probability axioms rules this out.[6]

Of course, it would be intuitively unreasonable to have background beliefs that make the one-burglar explanation superior while also giving it a lower prior conditional credence. But that does not change the fact that it is *possible* to be unreasonable in this way without violating the requirements of subjective Bayesianism, and that IBE and Subjectivist Conditionalization will disagree should such a possibility arise. Nor can we dismiss this possibility, and say that the subjective Bayesian can just patch the problem by adding requirements that will ensure coordination between background beliefs that make an explanation better and the explanation's prior conditional credence. To add such requirements is to turn to objective Bayesianism in response to van Fraassen's argument. In my view, this is exactly how the compatibilist should respond to van Fraassen's argument, but it offers no consolation to the compatibilist who sought to defuse van Fraassen's argument by demonstrating compatibility with Subjectivist Conditionalization. To fully appreciate this, notice that if we add this sort of constraint to the usual

6 Or Reflection, or the Principal Principle, or Regularity, or any other standard subjective Bayesian requirement.

subjectivist list of requirements, and somebody finds themselves in violation of it, they must choose between favoring the best explanation and obeying Subjectivist Conditionalization.

Second Objection: IBE tracks beliefs about explanatory goodness, not actual explanatory goodness. There are two ways to understand IBE, what we might call the "internalist" and the "externalist" readings. On the internalist reading, the explanation you should prefer is the one that *you judge to be* better, whereas the externalist reading enjoins you to prefer the the one that *in fact is* better.[7] One might think that it is only the externalist reading that is in danger of conflicting with Subjectivist Conditionalization, since the internalist reading makes your posterior credence dependent on the idiosyncrasies of your beliefs just like Subjectivist Conditionalization does. But the point made in response to the first objection applies here too. While the internalist reading might make your posterior credence depend on the idiosyncrasies of your prior credences, it makes it depend on a different idiosyncrasy than the one that matters to Subjectivist Conditionalization. It is possible to be very confident that H is a better explanation of E than H' is, and yet have $p(H|E) < p(H'|E)$; this does not contradict the probability axioms or any other standard Bayesian principle. So, whether IBE tracks real explanatory goodness or merely believed explanatory goodness, Subjectivist Conditionalization tracks conditional belief, which is something different.[8]

Third Objection: the best explanation just is the one with the highest conditional prior. I claim that that the one-burglar explanation is the better one even if your prior conditional degree of belief is higher for the two-burglar explanation. The compatibilist might simply insist that this isn't so, and that any appearance to the contrary comes from our being in the habit of having a higher prior conditional degree of belief for explanations like the one-burglar explanation. To reinforce this claim, she might point out that, if the two-burglar explanation has a higher prior conditional probability, then Bayes's theorem tells us that it has a higher prior probability and/or likelihood. That is, if $p(H|E) > p(H'|E)$, then $p(H)p(E|H) > p(H')p(E|H')$. And this might be interpreted as saying that H strikes a better

7 Notice, this distinction is orthogonal to the relativity of explanatory superiority to background belief. In the previous paragraph, we saw that different background beliefs affect which explanation is *in fact* better. But whichever one is in fact better, given your background beliefs, it is conceivable that you would mistakenly believe the other is superior.

8 We could add to the standard subjective Bayesian constraints a constraint that ensures coordination between conditional degree of belief and belief in explanatory goodness. But the spirit of IBE would also oblige us to add constraints coordinating beliefs about explanatory goodness with beliefs about best-makers — simplicity, unification, and so on — as well as coordination between beliefs about best-makers and beliefs about the features that really determine them — number of free parameters and the like. And this again is to save compatibilism by going objectivist, since we add substantial constraints on what prior degrees of belief are reasonable.

balance between fit with background belief — as captured in $p(H)$ — and fit to the data — as captured in $p(E|H)$. Thus H really is the better explanation if $p(H|E) > p(H'|E)$.

Of course, one can always rig one's account of explanatory goodness so that compatibility is guaranteed. If "best" just means "has the highest prior conditional credence", then compatibility with Subjectivist Conditionalization is a no-brainer. The compatibilist who gives the above account of explanatory goodness certainly is not guilty of anything so egregious. She identifies commonly-cited explanatory virtues with probabilistic ones — fit to background belief with $p(H)$, and fit to the data with $p(E|H)$ — and then relies on Bayes's theorem to bind explanatory goodness to prior conditional credence. But the question remains whether the notion of goodness she offers is thick enough to preserve the spirit of IBE.

I think the answer is clearly negative. On her account, fit with background belief and fit to the data are the only contributors to explanatory goodness, leaving out simplicity, elegance, unification, and the rest of the standard explanatory virtues. In fact, there is nothing intrinsic to the hypothesis or its relationship to the evidence that is relevant to explanatory goodness, on her picture. Explanatory goodness is, for her, *entirely* a matter of the agent's degrees of belief. This is surely not the way we ordinarily think about explanatory goodness, and this was the point of my above remark that what makes the two-burglar explanation a poor one is the intricacy of the narrative, the multiplicity of characters, entries, exists, and interactions. To understand "best" in the way we are now considering, we must depart from the criteria that we apply in such cases, whatever they may be. And when we do, it becomes very dubious whether we are respecting the spirit of IBE at all.[9]

If this does not discourage the compatibilist from understanding explanatory goodness entirely in terms of $p(H)$ and $p(E|H)$, I can only urge that she consider cases where those terms are undefined because the relevant prior degrees of belief do not exist. We will consider such cases in the next section.

3.2 The Heuristic View

I expect that many compatibilists will happily agree that explanatory goodness and prior conditional credence *can* diverge, but they will claim that they typically do not diverge. Compatibilists like Okasha (2000), McGrew (2003), and Lipton (2004)

9 We noted earlier that explanatory goodness must be partly relative to background assumptions, but the point is that it cannot be entirely a matter of the agent's degrees of belief. We also saw that IBE might be interpreted as a policy of preferring the explanation we *judge* or *think* to be better, rather than the one that actually is better. But that is different from saying that our degrees of belief are what *make* one explanation better than another, to the exclusion of all other factors.

stress IBE's value as a heuristic guide, helping us to meet ideal Bayesian standards. Okasha, for example, says that our explanationist inferences use "explanatory considerations as an aid to calculating the priors and likelihoods needed to apply Bayes's theorem," (Okasha 2000: p. 703) and Lipton and McGrew make similar pronouncements. I take it one motivation for this heuristic view of IBE is that it gives explanationist thinking a substantive role in an epistemology whose standards of correctness are ultimately Bayesian. But another motivation is, very likely, that treating IBE as a heuristic allows us to preserve it in the face of possible conflicts with Bayesian standards of correctness. If explanatory goodness is fallibly but reliably correlated with prior conditional credence, then it is plausible that following IBE is generally legitimate as a way of trying to meet Bayesian standards.

So the compatibilist's view on the break-in example will probably be this: it is possible for someone to have degrees of belief such that the worse explanation has the higher prior conditional credence. But such cases are not the norm, and because they are not the norm, IBE is a reliable guide to Subjectivist Conditionalization. Moreover, given how hard Subjectivist Conditionalization is for us to follow because of our cognitive limitations, we ought to use IBE as a heuristic to help us adhere to Subjectivist Conditionalization as best we can.

To the extent that I find Subjectivist Conditionalization plausible as a deep standard of rationality, this seems to me a reasonable and attractive position. But I do not find Subjectivist Conditionalization very plausible. And, I think that if explanationists attend to those cases where it is not plausible, they will be convinced that they concede too much when they demote IBE to mere heuristic status, and grant normative primacy to Subjectivist Conditionalization.

One of the famous limitations of Subjectivist Conditionalization is that it only applies when the requisite prior degrees of belief exist, while the history of science provides many examples where they do not. Major scientific breakthroughs provide striking examples, introducing wholly new concepts and theories that no one could have had a prior degree of belief in. But more mundane examples abound too. Small scientific breakthroughs, and even run of the mill research, can uncover hypotheses that no scientist could claim to have had well-defined prior degrees of belief in. Even just day to day experience provides hypotheses for which we do not have prior degrees of belief. I am right now wondering why I feel fatigued despite having drunk four cups of coffee. I think it most likely that the regular and decaffeinated pots have gotten mixed up, so that I have been drinking decaffeinated coffee all morning, but I had no prior degree of belief in that hypothesis when I walked into the cafe.[10]

10 I could have constructed such a degree of belief on the fly if I had been asked, upon entering the cafe, "how likely do you think it is that the pots have been switched, on the assumption that you will still feel tired after having had four cups?" But being able to construct a degree of belief is not

One of the chief advantages IBE has over Subjectivist Conditionalization is that it provides some basis for preferring one theory over another in such cases. Subjectivist Conditionalization has trouble accounting for the debates that swirl around transitions from one scientific theory to another, because it presumes grounds for such debates that do not exist. But IBE does not face this problem. While it is not always clear which theory is simpler, more unifying, more elegant, etc., it at least makes sense to debate the issue, and in some cases to declare one or another theory the winner. And to the extent that it is not clear which explanation is better, standoffs may be inevitable and reasonable. And the same goes for mundane cases like the coffee example. Much can be said in favor of the switched-pot hypothesis, while the alternative explanation that my body has suddenly lost its ability to metabolize caffeine as the result of some novel genetic mutation is clearly inferior.

It is especially noteworthy that each of Okasha, Lipton, and McGrew intends for IBE to apply in cases where Subjectivist Conditionalization does not apply because the requisite priors do not exist. Okasha, for example, is explicit that "IBE illuminates where Bayesianism is silent":

> [...] notoriously, Bayesians have nothing to say about situations where agents invent new hypotheses in response to experience. Change of opinion of this sort eludes Bayesian representation entirely [...] In those cases where agents respond to new evidence by inventing new hypotheses, the Bayesian model is silent. But IBE provides a useful, if schematic account of what is going on: the agents are trying to explain the new evidence... (Okasha 2000: pp. 706-7)

If the aim is to preserve IBE in a form that will let it illuminate where Bayesianism is silent, it is hard to see how the heuristic view will satisfy. On that view, IBE's normative force is entirely derivative on the correctness of Subjectivist Conditionalization. In cases where Subjectivist Conditionalization does not apply, we have no obligation to live up to it, and hence no reason to use IBE as a heuristic to do so. That is not to say that using IBE would be irrational, just that it is an arbitrary choice of one possible epistemic policy among many. Okasha could say that, when Subjectivist Conditionalization does not apply, IBE's normative force is no longer

---

the same as having one.

It is tempting to just define the degree of belief you have as the one that you would construct if asked to, but there will often be no fact of the matter. And unless we are prepared to say that I have a degree of belief in every proposition I might ever come to consider — *in the sense of 'degree of belief' that is relevant to Bayesian epistemology* — this is not a satisfactory conception of degrees of belief.

derivative on Subjectivist Conditionalization and is instead autonomous, but this is both ad hoc and implausible.

Lipton is not so explicit that IBE should apply where Bayesianism is silent, but his favored exemplar of IBE is a case where Bayesianism is silent. Lipton's example is Semmelweis's realization that childbed fever was caused by the transmission of "cadaverous matter" from the hands of attending surgeons and students to patients in labor. Semmelweis was the house officer of the First Obstetrical Clinic at the Vienna General hospital in 1846, where the mortality rate from childbed fever was approximately 13%, as compared with 2% in the Second Obstetrical Clinic. After considering and eliminating several potential explanations, Semmelweis noticed that a friend of his died of similar symptoms when he pricked his finger during an autopsy. Semmelweis conjectured that there was a connection between the handling of cadavers and the infections in the First Clinic, and found that physicians and students would frequently go straight to the First Clinic from the autopsy room, but not so with the Second Clinic. He concluded that infectious material was being transmitted from the cadavers to the patients in the First Clinic, but not the Second.

Semmelweis came to this explanation only after conceiving and abandoning several others, and after having the cadaverous matter hypothesis strongly suggested to him by the death of his friend. Plausibly, it had never occurred to him, and before his friend's death he had no prior degree of belief in either the evidence acquired or the hypothesis he ultimately settled on. And that is certainly true of those to whom he brought his results. Nevertheless, Semmelweis could readily appreciate the strength of the evidential support for his hypothesis, and it was on the force of that same strength that he ultimately convinced others. If Semmelweis's case is an exemplar of IBE, then IBE is not subservient to Subjectivist Conditionalization.

McGrew, like Lipton, is not explicit about whether IBE should illuminate where Bayesianism is silent. But he endorses an example that speaks for him:

> In 1979 two quasar images only five arcseconds apart, QSO 0957+561, were found to have identical spectral characteristics. Data on the spectra of known quasars indicated that there was only a remote probability of such a coincidence on chance; an explanation seemed called for. By far the most attractive hypothesis proposed was that the phenomenon consisted of a double image produced when radiation streaming from a single quasar was bent by the gravitational field of some massive objected located between us and the quasar — a gravitational lens. Pursuing this hypothesis, astronomers subsequently discovered a cluster of galaxies in the proper place to do the relativistic bending. (McGrew 2003: p. 563)

McGrew claims that the initial preference for the gravitational-lensing explanation is an instance of IBE. For the scientists involved, it may well be plausible that they had prior conditional degrees of belief that would make the normative force of that inference derivative on Subjectivist Conditionalization. But for me, that is not plausible. I can appreciate the elegance of the explanation, and its superiority to the alternatives, but only with a good deal of helping explanation from my friends who know the physics. My inclination to agree with the scientists involved that the lensing explanation was best cannot be consonant with Subjectivist Conditionalization, because I had no prior degrees of belief involving either the evidence or the hypothesis. McGrew might well say that my use of IBE here differs from the scientists's — that it does not inherit any normative force from Subjectivist Conditionalization, because Subjectivist Conditionalization does not apply for me. But this amounts to saying that there is a fundamental difference in the force of the reasoning of those who were on the inside before the research went down, as opposed to the rest of us who come into the know after the fact.

Defenders of the heuristic take on IBE might make the following response to this objection: even though IBE's normative force is ultimately derivative on the correctness of Subjectivist Conditionalization, that force does not just go away in cases where Subjectivist Conditionalization does not apply. IBE gets its force from its *general* agreement with Subjectivist Conditionalization, so it is as a *general rule* that IBE has that force. Even if this response preserves the normative force of IBE in some cases where Subjectivist Conditionalization does not apply, I do not see how it could preserve it in all cases, especially reflective cases. Suppose a case where we know that IBE is a heuristic approximation to Subjectivist Conditionalization, and we know that Subjectivist Conditionalization does not apply, as would be the case for me in McGrew's gravitational lensing example. If I know that IBE's correctness derives entirely from its use as an approximation to Subjectivist Conditionalization, and I know that it cannot serve that purpose in this case, what could possibly make me beholden to it? Perhaps the details of some account of epistemic normativity could vindicate the compatibilist here, but it is a tenuous possibility to hang our compatibilist hopes on.

As a final remark against the heuristic interpretation of IBE, notice that it doesn't just rob IBE of some of its most interesting applications; it also robs it of much of its intuitive appeal, since IBE inherits the extreme subjectivity of subjective Bayesianism. Return to the burglar example. A full-blooded explanationist will insist that, even if your prior conditional credence in the second explanation is higher, the first explanation is the one you should ultimately prefer. Assuming that your background beliefs are typical (with the exception of your conditional degrees of belief in the two explanations), it is just plain unreasonable to prefer the story about two burglars converging on your home on the same day, at the

same time, before being scared off by a corrupt and opportunistic police officer. Common sense dictates this. Were we to debate about which explanation to prefer, I could offer you several reasons to support the simpler story, and you could offer me none to prefer the baroque one. One of the best things IBE has going for it is that it agrees with common sense and our ordinary practices of reasoning and argumentation in cases like this. And the fact that it completely disregards all these considerations in cases like the burglar example is one of the worst things Subjectivist Conditionalization has going against it. Faced with such a conflict between IBE and Subjectivist Conditionalization, if the explanationist yields to the Bayesian, she spurns common sense and strong intuitions, undermining much of the attraction of her view.

I have been trying to cajole explanationists out of a heuristic view of IBE, by stressing how much of the interest and appeal of IBE is lost when we demote it to heuristic status. First, I pointed out that we sacrifice one of its most interesting applications, one that avowed compatibilists are committed to preserving (whether intentionally like Okasha, or unintentionally like Lipton and McGrew). Second, I pointed out that we rob IBE of much of its intuitive appeal.

I realize that these points will not move any committed subjective Bayesians to give up Subjectivist Conditionalization and embrace IBE. As criticisms of subjective Bayesianism, these points offer nothing new; those who endorse Subjectivist Conditionalization these days do so knowingly, despite the limitations and counter-intuitiveness I have been stressing. What I have been trying to show is that compatibility with subjective Bayesianism infects IBE with the same limitations and counter-intuitiveness. My hope is that, having seen this, explanationists who are not already committed subjective Bayesians will appreciate the limitations of compatibilism. I hope that their explanationist inclinations will then compel them to reject Subjectivist Conditionalization, and to embrace a more full-blooded understanding of IBE.

I suspect that the explanationist has much to offer the Bayesian, but the potential benefits of a marriage of the two views are lost when we make IBE subservient to Subjectivist Conditionalization. If we grant IBE primacy instead, and use it to shape a more objective Bayesianism that rejects Subjectivist Conditionalization, we are in a position to develop a Bayesianism that is free of the limitations that come with Subjectivist Conditionalization. Such a view would be a truer friendship between the explanationist and the Bayesian than the heuristic approach to compatibilism, since each party would stand to benefit from the other. The explanationist can help the Bayesian fill in the gaps in her view, by providing a way to construct priors for new theories, for example. And the Bayesian provides the explanationist with a powerful and well-studied formal framework in which to formulate the details of her view. The remainder of the paper is dedicated to

sketching the rough contours of this kind of Bayesian IBE, and outlining some suggestions for its further development.

4 IBE and Objective Bayesianism

Subjectivist Conditionalization uses prior degrees of belief as the probabilities to conditionalize on evidence, but there is an alternate Bayesian tradition that looks elsewhere for prior probabilities. Objective Bayesians like Carnap (1950) and Jaynes (1968) famously hold that the correct prior probabilities to use are determined by rules like the Principle of Indifference, rather than by whatever degrees of belief you happen to have just before the evidence comes in. The current proposal is that the explanationist should adopt the objectivist framework, and use explanationist thinking to fix objective prior probabilities, either in conjunction with, or in place of, existing objectivist principles. Ideally, explanationist considerations would complement existing objectivist principles like the Principle of Indifference, yielding a more applicable and reasonable brand of Bayesianism.

In sketch, the view goes like this: explanatory considerations and other objective Bayesian principles fix an objectively correct distribution of "a priori" probabilities, p, which describes the degrees of belief an agent with no evidence whatsoever ought to have. We could allow a range of reasonable prior distributions instead of just one but, for simplicity, let's just assume that a unique distribution would be determined by the completed account. Instead of Subjectivist Conditionalization, the objectivist endorses what we might call

Objectivist Conditionalization At any given time, your credence in an arbitrary proposition H ought to be $p(H|E)$, where p is the correct a priori probability distribution, and E is your total evidence at that time.

According to Objectivist Conditionalization, the appropriate posterior credences are determined by what p says about how your evidence bears on a hypothesis, rather than by whatever your prior credences happened to be.

The substance of an objective Bayesian theory lies almost entirely in the shape of the a priori probability function, p. The current proposal is that the explanationist should see her project of spelling out the details of IBE as part of the objective Bayesian's project of characterizing p. If we constrain p such that, whenever H is a better explanation of E than H' is in light of background assumptions B, we have

$$p(H|E \wedge B) > p(H'|E \wedge B),$$

then IBE and objective Bayesianism will be genuinely compatible. So, as the explanationist fills in the details of an account of explanatory goodness, she effectively

provides the objective Bayesian with constraints on p of the above sort. Ideally, these constraints would complement the work that objective Bayesians have already done to specify p, and maybe even solve some of the problems that have hampered the objective Bayesian program.

Regrettably, I have no detailed proposals to offer in this regard. But I do want to convince explanationists that they should pursue compatibility with objective Bayesianism rather than subjective Bayesianism, and also convince the objective Bayesians that they should welcome help from explanationists, so I must provide at least some concrete illustration of how this cooperative project might go. To that end, I will briefly suggest three ways that the explanationist might begin to cooperate with the objective Bayesian. Hopefully, these preliminary suggestions will lend some promise to the view that objective Bayesianism is a better fit for IBE than subjective Bayesianism.

Our first suggestion is the most simple-minded one: just adapt existing explanationist treatments of problems in confirmation theory, translating them into objective Bayesian terms. To illustrate, consider an explanationist treatment of Goodman's riddle suggested by White (2005). Goodman (1954) pointed out that our evidence that all observed emeralds are green can be subsumed under either of two generalizations, that all emeralds are green or that all emeralds are *grue*, where an object is grue just in case it is green and observed or blue and not observed.[11] Why should we believe the all-green generalization instead of the all-grue generalization? White's suggestion is that the all-green hypothesis offers a superior explanation of our data because it is more *stable*, where an explanation of a fact is stable to the extent that, according to that explanation, that fact couldn't easily have failed to obtain. While the all-grue hypothesis might explain why all the observed emeralds are green, its ability to explain is highly sensitive to something that could easily have been different, namely which emeralds we happened to observe. Assuming that the grue hypothesis is true, had we observed one of the emeralds that we in fact did not, it would have been bleen, not grue, a fact the all-grue hypothesis would not have been able to explain (because it would have been contradicted).

For the sake of illustration, let us take White to be right here. Our present goal is to use this kind of explanationist thinking to constrain a priori probabilities, and a straightforward translation of White's result gives us the constraint

$$p(\text{All Green}|\text{All Observed Green} \wedge B) > p(\text{All Grue}|\text{All Observed Green} \wedge B),$$

where B is any body of background information that makes the all-green explanation the more stable one. We don't always need explanationist treatments of

11 I follow White in simplifying Goodman's standard definition, which used 'observed by some specified future time t' instead of just 'observed'.

Goodman-type riddles to motivate such constraints, of course. We could just demand them for their self-evident sensibleness. But what the explanationist treatment offers us here is a general and fairly precise criterion of explanatory goodness — stability — which we can then use to lay down similar, more general constraints on p. Whenever H' is less stable than H as an explanation of E in light of B, we now know to demand $p(H|E \wedge B) > p(H'|E \wedge B)$, at least ceteris paribus. What the explanationist offers the objective Bayesian is a general criterion for constraining the a priori probabilities corresponding to inductive projections.

White's claim, that the all-green hypothesis offers the superior explanation because it is more stable, is certainly open to challenge, and I do not mean to endorse or defend it here. I use it only to illustrate how explanatory considerations might be used to constrain a priori probabilities. The example demonstrates one kind of help the explanationist can offer the objective Bayesian: as the explanationist develops precise criteria of explanatory goodness — stability, for example — the objective Bayesian can read off constraints for p. So far the collaboration between the explanationist and the objective Bayesian is fairly one-sided, with the objective Bayesian essentially cribbing off of the explanationist. The next suggestion is similarly one-sided, though the objective Bayesian will contribute substantially more when we come to the third suggestion.

Our second suggestion borrows an idea from the best-systems account of natural laws. According to the best-systems view, something is a law of nature just in case it is a theorem in all of the true deductive systems that best balance simplicity and informative strength (Lewis 1973). A classic objection to the view is that it robs laws of their explanatory power (Armstrong 1983). The fact that this particular emerald is green is, presumably, explained by the fact that it is a law that all emeralds are green. But if that generalization's status as a law comes to nothing more than its being true, and a part of the simplest and most informative deductive systems, how can it explain a particular instance? One possible response is to appeal to the connections between explanation, simplicity, and unification. One way to explain, says the best-systems advocate, is to locate the explanandum in a simple and orderly overall picture. So when the Humean explains the color of this emerald by appeal to the general law, she explains by locating this particular, local matter of fact in a simple, unifying, and informative picture of all the particular, local matters of fact.

Whether or not one buys the best-systems account of laws, this exchange is suggestive of a connection that the explanationist might wish to exploit — between the overall "organizedness" of a possible world and its explanatory goodness. Possible worlds that do not admit of much systematization — those worlds where the best deductive systematization is actually quite poor — might be regarded as explanatorily recalcitrant, and thus a priori improbable according to the explana-

tionist. Constraints on a priori probabilities could thus be derived from judgments of comparative systematizability between worlds.

If the explanationist also takes on the Humean's view that the physical chances in a world are those probabilities that appear in the best systematization, she can get even stronger constraints on a priori probabilities. Suppose that we have a chancy systematization, T, and we want to evaluate its a priori probability conditional on some possible body of evidence, E. Bayes's theorem tells us that[12]

$$p(T|E) = p(T)\frac{p(E|T)}{p(E)}$$

Supposing that T specifies a chance for E, $p(E|T)$ will be fixed by a chance-credence principle like the Principal Principle. Given the Humean supervenience of chances, $p(T)$ will be constrained by its goodness as a systematization, i.e. by the extent to which worlds with that systematization are not explanatorily recalcitrant and admit of simple, unified description. These considerations may not yield a definite value for $p(T|E)$, but they may be enough to warrant informative comparisons between T and its competitors. When comparing T to some competitor T' in light of evidence E, the denominator in Bayes's theorem drops out, leaving us to compare $p(T)p(E|T)$ and $p(T')p(E|T')$. Assuming that T', like T, specifies a chance for E — they are competing theories, after all — the likelihood terms will be fixed, and $p(T)$ and $p(T')$ are all that remain to be determined. To the extent that the explanationist can constrain these values then, she is in a position to say useful things about a chance-theory's conditional a priori probability.

Our third suggestion connects explanationist thinking to principles more traditionally associated with the objective Bayesian tradition. Objectivists typically assign a priori probabilities according to the Principle of Indifference, which says that each possibility is equally likely a priori. The trouble is that the results depend on how the space of possibilities is characterized. In cases where the set of possibilities is naturally represented in terms of a real-valued parameter, such as the time a train will arrive, the problem takes the form of alternative possible parameterizations, e.g. the speed at which the train travels. Which parameter we use to represent the set of possibilities affects what probabilities the Principle of Indifference assigns. Perhaps the explanationist can offer the objective Bayesian some help here.

Consider the following example of the parameter-dependence problem. Joel has to be in Montauk by 4 o'clock to meet Clementine, who will wait for him until

12 B is absent here because the probabilistic connections being discussed are purely a priori, holding in the absence of any background information. Alternatively, one can think of B as being contained in E.

4 and no later. He boards the 3 o'clock train and, while waiting for it to depart, anxiously asks the ticket collector what time they will arrive in Montauk. The ticket collector informs him that the train will arrive between 3:50 and 4:10, but refuses to be more specific. How confident should Joel be that he will arrive by 4 o'clock? The PI would seem to suggest that his credence should be 1/2, since a uniform distribution over the twenty minute interval between 3:50 and 4:10 assigns a probability of 1/2 to his arriving by 4. But then Joe thinks to himself: the ticket collector's information is equivalent to the train traveling at a speed between 51.4 and 72 mph, since it is a 60 mile trip to Montauk. However, a uniform distribution over the interval between 51.4 and 72 would yield a probability greater than 1/2 that the train's average speed will be $\geq$ 60 mph, i.e. that he will arrive by 4 o'clock. Which distribution should Joel use: a uniform distribution over arrival-time or over travel-speed?

There are actually infinitely many parameters Joel might choose from, not just travel-speed and arrival time. For example, if the interval $0 \leq x \leq 20$ represents arrival time, then any 1-1 map to another interval provides an alternative possible parameterization, $f(x) = x^2$ for example. And the infinite variety of possible parameterizations makes for an infinite variety of candidate probabilities for the outcome. A solution would be some reason to prefer one parameterization over the others, and this the explanationist may be able to provide. If Joel is an explanationist, he might approach the problem this way: he has been given the sparse information that the train will arrive in a certain time interval, and the best explanation for the constraints he has been given is that the train travels at a speed somewhere between 51.4 and 72 mph. Beyond that, each possible speed provides an equally good explanation of the constraints given to him, and so should be regarded as equally probable in the absence of further considerations. Thus Joel's degree of belief that he will make it in time to meet Clementine should be greater than 1/2 — about .58 to be more precise.

If that answer strikes us as incorrect, the explanationist may be able to accommodate us. The explanationist will say that it seems incorrect, if it does, because Joel has actually failed to hit upon the best explanation of his evidence. Joel supposes that the train travels at a constant speed, which seems unlikely on its own, and also fails to account for several features of his data about the arrival time: (i) the data was reported by the conductor, (ii) the conductor refused to be more specific, and (iii) the range of times is suspiciously symmetric around 4 o'clock. So perhaps the better explanation of his information is that the train's speed varies depending on the conditions on the tracks, so sometimes the train arrives a little early and sometimes it arrives a little late. Or maybe the train leaves a little earlier some days, a little later others. Or maybe both are true.

As potential explanations of the constraints multiply, it might seem that the

problem of parameter-dependence is right back with us. To some extent this is true; different explanations will suggest different parameterizations, and each parameterization will yield its own probabilities. But the under-determination here is not as bad as before. The vast majority of the infinity of possible parameterizations will be ruled out by their inability to explain the conductor's constraints. Also, the under-determination that remains is both appropriate and manageable. Joel should take the various possible explanations seriously, and hence should not calculate his probabilities on the assumption of any one of them. Instead, it seems reasonable that he should use a weighted average of the various explanations, weighted according to each explanation's goodness. For the sake of illustration, let us suppose that the only three explanations Joel must take seriously are that the departure-time varies, that the train's average speed varies, or some combination of the two. Let us also suppose that that is the order of explanatory goodness, from worst to best. Then Joel's appropriate credence is calculated by taking a uniform distribution over the explanatory parameter in each explanation, and then averaging the three results weighted according to the goodness of the explanations from which they are derived respectively.

The example thus illustrates a three-step algorithm that the explanationist offers for resolving the ambiguity in the Principle of Indifference:

i. Identify the potential explanations of the constraints given in the problem and their accompanying parameters.

ii. For each parameter, calculate probabilities according to a uniform distribution.

iii. Average the results, weighting each one according to the goodness of its accompanying explanation.

This explanationist solution is questionable in its rationale, and probably limited in application. The reasons for a uniform distribution over an explanatory parameter are obscure; indeed, the very idea of explaining constraints on one parameter in terms of constraints on another may itself be problematic. And in many cases, steps 1 and 3 will be inapplicable. Step 1 requires each explanation of the given parameter-constraints to have an "accompanying parameter", and it may not always be clear what parameter "accompanies" the explanation, if any. And step 3 will not be applicable when the various possible explanations are incomparable with respect to goodness. But despite its limitations and underdeveloped state, the proposal serves its illustrative purpose: it demonstrates how explanationism might be fit into an objective Bayesian framework, helping to constrain a priori probabilities. In a situation where B, H, and E are such that the Principle of Indifference applies but

does not yield a unique value for $p(H|E \wedge B)$, we can use the three-step algorithm just outlined to calculate (or maybe just further constrain) $p(H|E \wedge B)$.

5 Summary

I have been arguing that the compatibility of IBE and Bayesianism is not so simple a matter as one might think. Van Fraassen's argument that IBE conflicts with Subjectivist Conditionalization demonstrates a real conflict between the two rules, since they track different things. IBE tracks explanatory goodness or beliefs about explanatory goodness, while Subjectivist Conditionalization tracks prior conditional degree of belief. Moreover, resolving the conflict by demoting IBE to heuristic status robs IBE of much of its appeal and interest. IBE not only inherits the subjectivism of subjective Bayesianism, but also becomes inapplicable in many cases where explanationists and avowed compatibilists want it to apply. In response, I urged explanationists and compatibilists to abandon compatibility with Subjectivist Conditionalization, and embrace the objective Bayesian framework instead. To bolster the appeal and promise of that approach, I offered some preliminary suggestions for fitting IBE and objective Bayesianism together.

References

Armstrong, David. 1983. *What is a Law of Nature*. Cambridge University Press.

Carnap, Rudolf. 1950. *Logical Foundations of Probability*. University of Chicago Press.

Christensen, David. 2004. *Putting Logic in its Place*. Oxford University Press.

Douven, Igor. 1999. Inference to the best explanation made coherent. *Philosophy of Science* 66.

Goodman, Nelson. 1954. *Fact, Fiction, and Forecast*. The Athlone Press.

Greaves, Hilary & David Wallace. 2006. Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind* 115: 607–632.

Jaynes, Edwin T. 1968. Prior probabilities. *IEEE Transactions On Systems and Cybernetics* SSC-4(3): 227–241.

Jeffrey, Richard. 1970. Dracula meets wolfman: Acceptance vs. partial belief. In Marshall Swain (ed.) *Induction, Acceptance, and Rational Belief*. Reidel: Dordrecht.

Jeffrey, Richard. 1983. Bayesianism with a human face. *Minnesota Studies in the Philosophy of Science* X.

Lange, Marc. 1999. Calibration and the epistemological role of bayesian conditionalization. *Journal of Philosophy* XCVI.

Levi, Isaac. 1980. *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. The MIT Press.

Lewis, David. 1973. *Counterfactuals*. Harvard University of Press.

Lewis, David. 1980. A subjectivist's guide to objective chance. In Richard C. Jeffrey (ed.) *Studies in Inductive Logic and Probability*, vol. II. University of California Press.

Lipton, Peter. 2004. *Inference to the Best Explanation (Second Edition)*. Routledge.

Maher, Patrick. 1993. *Betting on Theories*. Cambridge University Press.

McGrew, Timothy. 2003. Confirmation, heuristics, and explanatory reasoning. *British Journal for the Philosophy of Science* 54.

Okasha, Samir. 2000. Van fraassen's critique of inference to the best explanation. *Studies in the History and Philosophy of Science* 31(4).

Teller, Paul. 1973. Conditionalization and observation. *Synthese* 26.

Tregear, Mark. 2004. Utilising explanatory factors in induction. *British Journal for the Philosophy of Science* 55.

van Fraassen, Bas. 1984. Belief and the will. *The Journal of Philosophy* 81(5): 235–256.

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

van Fraassen, Bas. 1995. Belief and the problem of ulysses and the sirens. *Philosophical Studies* 77: 7–37.

van Fraassen, Bas. 1999. Conditionalization: a new argument for. *Topoi* 18.

White, Roger. 2005. Explanation as a guide to induction. *Philosopher's Imprint* 5(2).