

Against Explanationist Bayesianism

Lok C. Chan

1 Introduction

The crux of the problem here ultimately boils down to the status of what I call the compatibilist thesis:

Compatibilism: Inference to the best explanation is compatible with some form of Bayesianism.

Van Fraassen begins the debate by arguing that this statement is *false*, and this is countered by many explanationists.

For the sake of clarity, we must first decide what being a “Bayesian” entails. It certainly won’t do to call anyone who uses Bayes’ theorem a Bayesian, since the theorem itself is a logical consequence of the theory of probability, so anyone who accepts the axioms of probability is a “Bayesian” in that sense. We also ought to distinguish Bayesianism in epistemology with Bayesianism in statistics. Even the strictest Bayesian statistician, who might reject all uses of Frequentist methods, could consistently maintain that Bayes’ theorem is not the *only* rational way to revise one’s

belief in light of new evidence.

Since explanationists are responding to van Fraassen's claim that Bayesianism and explanationism is incompatible, I suggest that we ought to retain the terminology used by van Fraassen, especially since he also tends to be more careful about it than his opponents. To begin, he distinguishes *Bayesianism* from *probabilism*.¹ *Probabilism* can be characterized by the acceptance of the two statements below:²

1. There are *partial* beliefs, which are measurable in terms of degrees.
2. The axioms of probability govern the coherence of partial beliefs.

Probabilism is rather broad and unambitious position—all it asserts is that our judgments are *translatable* to formal propositions, which can be analyzed and criticized in terms of the theory of probability. Probabilism, as van Fraassen defines, does not even require, as a matter of descriptive fact, that people have numerically precise probabilities in their heads, but only that their judgments are analyzable as probabilities.³ This is no different than saying that our arguments can be systematically translated into symbolic languages such as propositional logic or predicate logic.

Probabilism is often not defined, or even mentioned, by explanationists who responded to van Fraassen's argument, but such an oversight is problematic, because it is easy to mistakenly assume that van Fraassen is arguing that IBE is incompatible with probabilism. Such a claim, however would require a much stronger argument than the one offered by him. He does argue, however, that IBE is incompatible with what he calls *Orthodox* Bayesianism.⁴ An Orthodox Bayesian, as describes by him, is

¹Bas C. van Fraassen, *Laws and Symmetry* (Oxford University Press, 1989), 175.

²Fraassen Ch 7 Sec 2.

³Fraassen, 154.

⁴Fraassen, 151.

someone who believes that a rational agent responds to new evidence by, and only by, revising her prior beliefs by following what Bayesians call conditionalization:

Conditionalization: one should revise her prior degree for the belief H in light of evidence E via the application of Bayes' theorem, which results in the posterior opinion degree of belief $P(H|E)$.

Orthodox Bayesianism, then, is a fairly restrictive thesis: how we update our opinion is a strictly logical endeavor. It is important to keep in mind that *this* is what van Fraassen has in mind when he says that Bayesianism and explanationism are incompatible.

In its most naive form, inference to the best explanation(IBE) says that we should infer to the hypothesis that best *explains* the evidence we have. IBE is originally proposed by Gilbert Harman as a replacement of enumerative induction, that is, the process of draw inference about a population as a whole from samples.⁵ We can call this *strong inference to the best explanation*:

SIBE: One is entitled to infer from the the fact that a hypothesis H would best explain the evidence E to the conclusion that H is true.

Note that in this context the term “inference” is used in the logical sense: the conclusion of the inference, that the best explanation is true, is guaranteed, or, as van Fraassen calls it, *rationally compelling*. Fraassen attacks SIBE relentlessly. This is no doubt related to the fact that strong inference to the best explanation is used to motivate scientific realism, a position van Fraassen attacks in his *The Scientific*

⁵Gilbert H. Harman, “The Inference to the Best Explanation,” *Philosophical Review* 74, no. 1 (1965): 90.

Image.⁶ Van Fraassen’s argument against SIBE, in the most general form, is that there is no such thing as the *best* explanation—just the best *out of the explanations available to us*. Van Fraassen argues this is a horrible justification for a belief, since for some reason we might only have horrible explanations available to us, so ‘our selection may well be the best of a bad lot.’⁷ In any case, while this is necessary background information, the status of SIBE or scientific realism is not central to our concern, since they pertain to *full* beliefs.

Assuming SIBE is not viable, van Fraassen asserts that the strongest recourse available the supporters of IBE is *entrenchment*, which amounts to repackaging IBE into a rule that works well with Bayesianism. The more plausible way to do this, according to van Fraassen, is to give explanatory virtues a place in the revision of belief in light of new evidence: ‘behind the naive rule of IBE there might lie a recipe for adjusting our personal probabilities, in response to new experience, under the aegis of explanatory success’. Let’s call this new rule ‘probabilistic inference to the best explanation’, which entitles us to raise the probability of the best explanation:

PIBE: One is entitled to infer from the the fact that a hypothesis H would best explain the evidence E to the conclusion that H is *more probable*.

Van Fraassen, however, argues that this cannot do. To begin, if IBE is to be harmonized with Bayesianism, since it must not clash the Bayesian procedure of belief revision, i.e., conditionalization, but van Fraassen insists that this cannot be done without violating the Bayesian standard of rationality. To begin, he suggests that the new IBE-inspired rule must be *ampliative*: it goes beyond what is logically implied by

⁶Bas C. Van Fraassen, *The Scientific Image* (Oxford University Press, 1980), 19–21.

⁷Fraassen, 143.

our evidence. In contrast, Bayes' theorem is *explicative*, since the posterior probability is nothing but an arithmetic consequence of conditional probability. This must mean that the PIBE is the rule that confers 'bonus' probability to a belief based on its explanatory virtue. This is where PIBE conflicts with Bayesianism.

A simplified version of his argument is as follows. Suppose we are interested in the bias of a certain coin, θ , which indicates the probability of the coin landing on heads: $\theta = 1$ if it is certain to land on heads, and 0 if certain to lands on tails. Suppose we know that there are 3 equally probable hypotheses: (A) $\theta = 0.9$, (B) $\theta = 0.5$, and (C) $\theta = 0.1$. Suppose our evidence gathering process is described as follows: $X_i = 1$ denotes 'the coin has landed on heads on the i toss' and $X_i = 0$ otherwise. Suppose we have toss the coin 4 times, and they all landed on heads. So the evidence E is $\sum_{i=1}^4 X_i = 4$. Now, we need to find out the overall probability of E :

$$\begin{aligned} P(E) &= P(A)P(E|A) + P(B)P(E|B) + P(C)P(E|C) \\ &= (1/3)0.9^4 + (1/3)0.5^4 + (1/3)0.1^4 \\ &= 0.24 \end{aligned}$$

Using Bayes' theorem, the posterior probabilities are: $P(A|E) = 0.9129$, $P(B|E) = 0.0869$, and $P(C|E) = 0.0001$. So far so good—4 consecutive heads favors the hypothesis that the coin is bias toward heads, which is what Bayes' thereom is showing us. What conditionalization suggests is that taking E as evidence, the rational way to revise our opinions is to reassign them in accordance to the posterior

probabilities.

Where does PIBE comes in, then? According to van Fraassen, PIBE's role is to add an extra step after conditionalization: giving bonus probability to the best explanation. Out of the three hypotheses, A —the coin is highly biased—best *explains* why we see nothing but heads, so PIBE entails the redistribution of the probabilities so that $P(A)$ would be even higher. Suppose we raise $P(A)$ to 0.999. This amounts to giving the best explanation a bonus of 0.086 in probability. To accommodate this, we can lower the probabilities of the other hypotheses accordingly. For instance: $P(A) = 0.999$, $P(B) = 0.00086$, and $P(C) = 0.00014$. So we have extrapolated ampliatively beyond what the evidence tells us by using PIBE.

Van Fraassen argues that this move is irrational, according to the standard of rationality accepted by Bayesians. In particular, he argues that PIBE violates the Bayesian notion of *coherence*, by pointing out that anyone who uses PIBE is subject to a *Dutch book*: a set of bets that ensures whoever takes these bets a net loss. Imagine that we are back at the beginning before we tossed 4 heads. Before tossing the coin for four times, we were offered the following set of bets. Let E again be 'the coin is tossed 4 times and they are all heads' and H be the $X_5 = 1$, that is, 'the fifth toss turns up heads'. Three bets are offered to us regarding our attitude toward the E and H :

1. \$10,000 if E and $\neg H$.
2. \$1300 if $\neg E$.
3. \$300 if E .

Now, using the standard *expected utility* model, the worth of each oh these bets

is explicated as the product of the prior probability of the winning condition, and its value:

1. Bet 1: $(10000)\frac{0.8^4(0.2)+0.5^5+0.2^4(0.8)}{3} = 323.16$
2. Bet 2: $(1300)(1 - 0.158) = 988.56$
3. Bet 3: $(300)0.158 = 71.87$

So these bets would be worth \$1383.6 in total. Suppose we bought these bets for exactly that much from a bookie, who then proceeded to toss the coin for 4 times. Either E is true or it is false, so there are two mutually exclusive way that the event could play out. The argument is that in both case we will lose money. Suppose E is false—at least one toss landed on tails. In this case, we would have won bet 2 but lost 1 and 3. We would receive \$1300 but this would still lead to a total loss of $-1383.6 + 1300 = -83.6$.

On the other hand, suppose E —all tosses turned up heads. We would receive \$300 by winning bet 3, and now bet 1 would depend entirely on the fifth toss. Now, van Fraassen asks, what should our degree of belief for $\neg H$, that the fifth toss will land on tails? Recall by PIBE we ought to give a bonus to the most explanatory hypothesis. In this case, it would be A , which says that the coin is highly biased. As before, we raise the pre-PIBE posterior $P(A|E)$ from 0.9129 to 0.999. At this point, the probability of tails would be

$$1 - P(H) = 1 - (0.999(0.9) + 0.00086(0.5) + 0.00014(0.1)) = 0.1$$

Suppose the bookie offers us \$1000 to buy bet 3 back. We would regard it as

fair and accept his offer, since the bet is now worth $10000(0.1) = 1000$. In this scenario, we end up with $-1383.6 + 300 + 1000 = -83.6$ —we incur exactly the same loss as we would if E were false. We have been “Dutchbooked”, which is a sign of irrationality according to Bayesianism. This happens exactly because we went beyond conditionalization with PIBE. Thus, van Fraassen concludes that compatibilism must be false: Bayesianism and Explanationism are inconsistent with each other.

Explanationists demur. There are two types of responses: the first and the popular tactic is to challenge the definition of PIBE involved: they accept van Fraassen’s Dutch book argument, but reject his formulation of PIBE. The second approach is to meet the Dutch book argument head on: accept van Fraassen’s formulation, but reject that it has been shown to be an irrational rule. The two approaches not only differ in the formulation of PIBE they accept, but also the sort of Bayesianism they assume. Let us examine them in turn.

2 Subjective Bayesianism and the Psychological View

Some explanationists find van Fraassen’s formulation of PIBE wanting. After all, it is natural for explanationists to be suspicious about how the opponent characterizes their view—what if there is an alternative formulation PIBE that circumvents van Fraassen’s Dutch Book argument? It might not be necessary that PIBE is about artificially raising the posterior probability of the best explanation. Alternatively, explanationists could argue for a role for explanatory virtues external to the formal

process of conditionalization. One route that some compatibilists have taken is to argue that explanatoriness has a role in how conditionalization is *realized* in human psychology, especially in light of how people arrive at their subjective priors. I shall call this the *psychological* view of compatibilism.

The psychological view is closely tied to the subjective variety of Bayesianism. To begin, note that Orthodox Bayesianism as defined earlier makes no claim about what prior probabilities come from, even though they are necessary for Bayes' theorem and, therefore, conditionalization. By itself, it's an incomplete epistemology: there needs to be an account about the rational status of prior opinions, and the standard of admissibility for conditionalization.

One view is that priors are *subjective*: before conditionalization, priors need not be justified, as long as they are consistent with the axioms of probability. Van Fraassen closely associates, if not identifies, Bayesianism with Subjective Bayesianism, for he considers Bayesianism, as an epistemological position, an attempt to address the philosophical problem of skepticism, which claims that knowledge is impossible, because we cannot justify all of our existing opinions. The subjectivist response is to reject the demand for justification for our existing opinions, but objectivity is maintained by a strict enforcement of conditionalization. The thought is that by following nothing but conditionalization, two opposite opinions will eventually converge to the same truth. Two Subjective Bayesians, no matter how much initial disagreement they might have about the bias of a coin, will both arrive at the same conclusion if they toss the coin for long enough. This effectively dislodges a powerful assumption of the skeptic.

More important, this also means that, to say that priors are subjective means something very specific in our epistemological context. When Bayesian *practitioners* speak of subjective priors in a decision-theoretical or statistical context, they usually just mean that for each problem, there is no “correct” parameters for prior distributions, but the lack of consensus should not deter one from using Bayesian methods about the said problem. This says nothing about what beliefs are rational. On the other hand, the picture of an ideal agent painted by what we call Subjective Bayesianism here is someone whose epistemological moves are to be underwritten by nothing but conditionalization. This is why van Fraassen describes a (Subjective) Bayesian as someone who regards any ampliative extrapolation—beyond what is entailed by experience and logic—as being irrational. The strict adherence to conditionalization is what allow a Subjective Bayesian to, as van Fraassen engagingly describes, “live a happy and useful life by conscientiously updating the opinions gained at his mother’s knees, in response to his own experience thereafter.”⁸

It is no wonder, then, that the extralogical bonus given by PIBE clashes with conditionalization. Some explanationists, such as Okasha(2000) and Lipton(2004), sees an opening in the idea that van Fraassen has unfairly stacked the cards against explanationism by asking it to do the impossible: providing extralogical support in the strictly logical procedure of conditionalization. The alternative, the argument goes, is to shift the locus from the mechanical procedure of conditionalization to the richer setting of how Bayesian reasoning is realized in human psychology.

One suggestion is that explanatory reasoning is how people come up with the numbers needed for Bayes’ theorem. Since Subjective Bayesianism does not legislate

⁸Fraassen, *Laws and Symmetry*, 1989, 178.

how priors ought to be determined, some explanationists think that this is where PIBE can fill the gap. This is specifically proposed by Samir Okasha.⁹ The thought is that when we compare two explanations, we often find the better explanation to be more probable, and that the better explanation should also “fit” the data better. This goes back to Gilbert Herman’s original proposal that a good explanation should have a high prior probability and/or a high likelihood—probability of the evidence conditional on the explanation.¹⁰ That is, take two hypotheses H_1 and H_2 , if H_1 is a better explanation of some evidence E than H_2 , then the following inequalities should hold:

$$P(H_1) > P(H_2)$$

$$P(E|H_1) > P(E|H_2)$$

So, the argument is that PIBE is not about giving bonuses to the posterior probability; instead, explanatoriness relates to probability by informing by how subjective priors and likelihoods are formed. How this works, presumably, is that human beings evaluate their personal probabilities by using intuitive explanatoriness as a shortcut.

This, however, is an empirical claim. Okasha primarily appeals to our intuition about this. Peter Lipton does use the findings in behavioral economics done by Kahneman and Tversky, such as the conjunction fallacy, to support the idea that people are using explanation as a heuristic to evaluate probability.¹¹ This, however, is another unsubstantiated empirical claim. While we can say for sure that people in

⁹“Van Fraassen’s Critique of Inference to the Best Explanation,” *Studies in History and Philosophy of Science Part A* 31, no. 4 (2000): 691–710.

¹⁰Gilbert Harman, “Knowledge, Inference, and Explanation,” *American Philosophical Quarterly* 5, no. 3 (1968): 169.

¹¹Peter Lipton, *Inference to the Best Explanation* (Routledge/Taylor; Francis Group, 2004) ch 9.

general do not think about probabilities like a Bayesian, explanatory reasoning is but one possible hypothesis. One could, for instance, just easily explain it by appealing to the representative heuristic.¹² Bayesians can also hitch their wagon to the effort to use Bayesian theory of confirmation to explain the conjunction fallacy.¹³

Putting the issue of empirical adequacy aside, there is also a conceptual issue. The proponents of the psychological view almost always appeal to contrastive cases with evidence already in hand: this hypothesis is a better explanation of the evidence than the alternative, so it must have a higher prior and/or a higher likelihood. Okasha, for instance, talks about an imaginary case in which a doctor evaluates the probabilities of two competing diagnoses by considering which explains the evidence—the symptoms—better. So the explanationist interpretation would be that the doctor is using how well each diagnosis explains the symptoms as a mental shortcut for calculating their respective posterior probability, just as how Kahneman and Tversky’s subjects use Linda’s description as evidence to be explained by the fact that she’s a feminist banker as opposed to a banker.

When we have existing facts in need of an account, I agree that people might find explanatoriness a convenient anchor for thinking about probabilities, but the problem is that this style of reasoning clashes with how conditionalization works. The only time when the prior is not under the constraint is when our opinion is entirely naive, that is, before receiving any evidence. This is when PIBE is supposed to intervene without subjecting one to the threat of a Dutch book. However, the type of PIBE

¹²Daniel Kahneman and Amos Tversky, “Subjective Probability: A Judgment of Representativeness,” *Cognitive Psychology* 3 (1972): 430–54.

¹³Vincenzo Crupi, Branden Fitelson, and Katya Tentori, “Probability, Confirmation, and the Conjunction Fallacy,” *Thinking and Reasoning* 14, no. 2 (2008): 182–99.

cases to which the explanationists appeal presupposes the existence of some facts to be explained, so it cannot offer any guidance here. Because of this, I must set a prior without the consulting any explanation. But PIBE cannot do anything after I get some evidence either: once I receive new evidence, I am supposed to carry out conditionalization. If I change my prior after seeing the new evidence, a Dutch book can be made against me, because I am arbitrarily redistributing my personal probabilities.

Proponents of the psychological view, I think, overlooks the importance of *ex ante*—before receiving evidence—evaluation of one’s prior, and the *ex post*—after the evidence—conditionalization. This is easily shown by using an example. Let’s say I am interested in the proportion of Duke students who are left-handed. Assume that I have never investigated this matter, so my opinion of it is based on nothing but my background knowledge about Duke and handedness. Since there is no evidence on which I can conditionalize my opinion, it would seem that this is the context in which IBE can shine, at least according to the proponent of the psychological view. However, my prior at this point in time cannot be informed by any explanatory consideration, because there is nothing to be explained—all I have is my subjective opinion about the proportion of left-handed students at Duke. Suppose I determine that the probability of randomly picking a left-handed Duke student, $P(L)$, is 0.2. So far my decision is *ex ante*. Suppose after ascertaining my prior, I then find 5 Duke students at random, and it turns out all 5 are left-handed. *Ex post*, it would appear that my initial prior is a poor explanation of the data, but, if I were a Subjective Bayesian, all I could do is to use Bayes’ theorem to generate my posterior probability. If I artificially adjust my prior to fit the data *ex post*, the result is no different than giving a bonus to the

posterior, because a bookie would have offered me the bets at the values I announced before the evidence. So, at the end, there is no room for PIBE here.

3 Flight from Orthodox Bayesianism

At this point, explanationists should wonder if they have been playing a rigged game: Subjective Bayesianism, as van Fraassen defines it, does not seem to leave *any* room for alternative ways of revising our belief in light of new evidence, due to its strict adherence to conditionalization. I think this o

Jonathan Weisberg, I think, comes to closest to

An Objective Bayesian, as defined by Weisberg, rejects Orthodox Bayesianism (which he calls Subjective Conditionalization) and instead adheres to the following:¹⁴

Objective 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 the time.

The difference between this principle and the subjective variant we have been working with is that

It is important to note, then, that the sort of objectivity Weisberg is suggesting is logical and not empirical.

Weisberg makes a few suggestions about how IBE can help determining this

¹⁴Jonathan Weisberg, “Locating Ibe in the Bayesian Framework,” *Synthese* 167, no. 1 (2009): 137.

objective probability function. He raises an example in which a person is trying to determine if the train is going to be late on a 60 mile trip. The train departs at 3 but she is supposed to have a meeting at 4, sharp. The conductor then inform her that the train will arrive between 3:50 and 4:10. Now, Weisberg asks, what should the agent's degree of belief that she will make it on time be? If we apply the Principle of Indifference over the time interval, it would seem that there is an equal chance that she will be late. However, Weisberg argues that she could just easily interpret the information as about the travel-speed of the train, which is translated into saying that the train is traveling between 51.4 and 72 miles per hour.

I do not think it is as straight-forward as Weisberg puts it, however.

Crupi, Vincenzo, Branden Fitelson, and Katya Tentori. “Probability, Confirmation, and the Conjunction Fallacy.” *Thinking and Reasoning* 14, no. 2 (2008): 182–99.

Fraassen, Bas C. Van. *The Scientific Image*. Oxford University Press, 1980.

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

Harman, Gilbert. “Knowledge, Inference, and Explanation.” *American Philosophical Quarterly* 5, no. 3 (1968): 164–73.

Harman, Gilbert H. “The Inference to the Best Explanation.” *Philosophical Review* 74, no. 1 (1965): 88–95.

Kahneman, Daniel, and Amos Tversky. “Subjective Probability: A Judgment of Representativeness.” *Cognitive Psychology* 3 (1972): 430–54.

Lipton, Peter. *Inference to the Best Explanation*. Routledge/Taylor; Francis Group, 2004.

Okasha, Samir. “Van Fraassen’s Critique of Inference to the Best Explanation.” *Studies in History and Philosophy of Science Part A* 31, no. 4 (2000): 691–710.

Weisberg, Jonathan. “Locating Ibe in the Bayesian Framework.” *Synthese* 167, no. 1 (2009): 125–43.