Against Explanationist Bayesianism

Lok C. Chan

1 Bayesianism, Explanationism, and Voluntarism

The positions I argue against In this dissertation, I always refer IBE as "IBE" and reserve "abduction" for choices and inferences made in service of setting up the environment in which induction can be carried out.

For an example of a modern use of the term 'abduction', consider Jerry Fodor: Inferences in which features of an embedding theory affect the inferential-cum-causal roles of their constituent beliefs are what philosophers sometimes call "global" or "abductive" or "holistic" or "inferences to the best explanation." From now on, I'll use these terms more or less interchangeably.[1] ## Probabilism and Bayesianism

In the philosophical context, the term "Bayesianism" is to be understood as an overarching epistemological position about the rationality of our beliefs. It certainly won't do to call anyone who uses Bayes' theorem a Bayesian, since the theorem itself is a logical consequence of conditional probability, so anyone who accepts the axioms of probability is a "Bayesian" in that sense. We also ought to distinguish Bayesianism in

---

[1] ???

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.[2]

I will begin by spelling out a foundational thesis that encompasses the types of views relevant to our discussion: *probabilism* is the acceptance of the two statements below:

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

Strict Bayesianism, as an epistemological thesis, is the position that accepts the tenets of probabilism, plus the thesis that *conditionalization* is only rational way to revise one's belief:

> *Conditionalization*: it is rational to update one's prior degree for belief H in light of evidence E via the application of Bayes' theorem, which determines the posterior opinion degree of belief $P(H|E)$.

Thus, Strict Bayesianism claims that that a rational agent is one whose belief can be captured by some prior distributions, which would be mechanically revised whenever data is introduced. In our particular context, the term "Bayesian" refers to someone who thinks that not only is this revision justifiable only by conditionalization, but also that rationality mandates it. A person does not have to a Bayesian to think that conditionalization can be just the right thing to do. One, for instance, may think

---

[2]This does not mean that philosophical Bayesianism and statistical Bayesianism are entirely separate—in fact, I think that Bayesian statistics is should be the model of a general Bayesian rationality. The position I advocate reflects this idea.

that while conditionalization is a perfectly fine principle, it cannot be the only rule, since there are occasions in which it is not applicable, e.g., a required prior might not be available.

In light of this, we can distinguish at least two ways in which we may say conditionalization is a rational process:

Persistent Rationality requires the *persistent* application of conditionalization—a rational agent always revise her initial opinions in light of evidence using Bayes' theorem.

Situational Rationality requires the *situational* application of conditionalization—a rational agent has to deliberate on the relevant facts to decide whether conditionalization is appropriate.

In our context, I take Strict Bayesianism to hold the requirement of conditionalization is persistent: it is not something that the rational just chooses to follow at will. S

Strict Bayesian, however, is incomplete: since conditionalization presupposes the existence of a prior distribution for the belief H supported by evidence E, Bayesians need an account for the status of these priors. In particular, it must address whether or not prior degrees of belief ought to be justified for it to be admissible to conditionalization.

A prominent view is that priors are *subjective*:

Subjective Bayesianism: priors need not be justified, as long as they are consistent with the axioms of probability.

For example, for any proposition H, $P(H)$ must be between 0 and 1, and $P(\neg H) = 1 - P(H)$. An early prominent of this view is F. P. Ramsey, who says:

> ...to ask what initial degrees of belief are justified... seems to me a meaningless question; and even if it had a meaning I do not see how it could be answered.[3]

From an epistemological point of view, Subjective Bayesianism can be considered as a type of anti-foundationalism. The idea is that the subjective view frees us from the burden of justifying our current opinions and holds that all rationality demands from us is to update our priors—whatever they are and however they were formed—by applying the rule of conditionalization diligently. This is an attractive feature of the subjective view, since it dislodges skepticism's most powerful premise, that rationality demands all our beliefs to be justified. This allows Bayesianism to be harmonized with the imagery of the Neurathian ship:

> We are like sailors who on the open sea must reconstruct their ship but are never able to start afresh from the bottom. Where a beam is taken away a new one must at once be put there, and for this the rest of the ship is used as support. In this way, by using the old beams and driftwood, the ship can be shaped entirely anew, but only by gradual reconstruction.[4]

In other words, rationality does not require us to build our beliefs from the ground up. We have to our initial opinions as given while we revise them with a rational procedure, which, according to Bayesians, *must* be conditionalization.

---

[3]Frank P. Ramsey, *Philosophical Papers* (Cambridge: Harvard University Press, 1990), 88.

[4]O. Neurath, Marie Neurath, and Robert S. Cohen, "Empiricism and Sociology," *British Journal for the Philosophy of Science* 25, no. 4 (1973): 199.

Objective Bayesianism: Only objectively justifiable priors are admissible.

In other words, there *are* rational degrees of beliefs, in addition to the subjective ones, and only rational ones are admissible to conditionalization. The earliest example of this is Laplace's "Principle of Non-Sufficient Reason":

> The theory of chances consists in reducing all events of the same kind to a certain number of equally possible cases, that is to say, to cases whose existence we are equally uncertain of, and in determining the number of cases favourable to the event whose probability is sought. The ratio of this number to that of all possible cases is the measure of this probability, which is thus only a fraction whose numerator is the number of favourable cases, and whose denominator is the number of all possible cases.[5]

In other words, this principle dictates that when we are in complete ignorance about the out of some event X, the probability *ought* to be equally distributed among all possibilities. For example, if in front of me is a coin I have never seen before in my life, I should believe that the probability of the coin landing on heads ought to be the same as the probability of it landing on tails, which is 1/2. Note that the principle says that in such a situation, the only justifiable prior is 1/2—any other degree would be irrational. The Principle of Non-sufficient Reason is not the only way for a degree of belief to be justified. Instead, an Objective Bayesian might hold that the only rational priors are the ones empirically extracted from the data. We shall return to this when we consider Jonathan Weisberg's view.

---

[5]Pierre-Simon Laplace, *Philosophical Essay on Probabilities* (Spring, 1995), 4.

1.1 Explanationism

The primary target of the first half of this chapters is *explanationism*, which holds that inference to the best explanation(IBE) as an indispensable rule of *inductive* inference. In its most naive form, IBE says that we should inferbelo the hypothesis that best *explains* the evidence we have. In our context, explanationism refers to the position that IBE is viable *inductive* rule of inference. In fact, in the original articulation by Harman, IBE is conceived to replace enumerative induction.

In this chapter, I will not concern myself with naive explanationism. Instead, my focus is on the explanationists who accept and argue for what I call the *compatibilist* thesis:

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

The compatibilists are almost universally motivated by van Fraassen's argument against a non-compatiblist version of Explanationism. Van Fraassen is well-known for arguing against IBE in its original, non-probabilistic, form. In its most naively powerful form—a view that van Fraassen does ascribe to some philosophers—IBE can be construed as a solution to Hume's problem of induction, which holds that there is no independent justification for extrapolating inductively beyond the evidence we have. IBE gets us out of this problems by giving justificatory force to explanatory virtues, so that the best explanation is the one we *should* accept. Van Fraassen attacks this position relentlessly. One often cited argument of his is that we never pick the best explanation *simplicter*, but 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.' (143)

Van Fraassen suggests that the strongest recourse available the supporters of IBE is *entrenchment*, which amounts to the 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' (PIBE), which entitles us to raise the probability of the best explanation.

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 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. 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 $\sum_{i=0}^{4} X_i = 4$. The marginal probability for E is:

$$P(E) = P(A)P(E|A) + P(B)P(E|B) + P(C)P(E|C) = 0.24$$

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 conditionalization is showing us. Where does PIBE comes in, then? Van Fraassen asserts that an argument from PIBE would be as follows: out of the three hypotheses, A best *explains* why we see nothing but heads: it's because it's highly biased. So what PIBE should do is to recommend 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.

This line of reasoning, however, flies in the face of the Bayesian notion of *coherence*, since it renders one subject to an so-called *Dutchbook*, which is a set of bets that ensures whoever takes these bets a 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'.

1. $10,000 if E and $\neg H$.

2. $1300 if $\neg E$.

3. $300 if E.

Now, we can calculate the values of these bets based on our prior probabilities:

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. Suppose it's 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 per 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 that we have used PIBE to give a bonus to the most explanatory hypothesis, A, which effectively has raised the marginal probability of H to 0.9. At this point, bet 1 is now worth $(10000)P(\neg H) = (10000)0.1 = 1000$. Suppose the bookie offers us exactly $1000 to buy this bet back. We would regard it as fair and accept his offer. 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.

2   Against Explanationist-Bayesian Compatibilism

In response to van Fraassen's argument, many arguments for compatibilism from the explanationist perspective have emerged. In particular, I will focus on Okasha (2000), Lipton (2004), and Weisberg(2009). There are important subtle differences in their views, but these explanationists do share a common strategy in pointing out that van Fraassen's characterization of PIBE is not mandatory, and that PIBE can be construed in a way that renders it not only compatible with, but also, so they argue, complementary to Bayesianism. Okasha's way is to hold that what PIBE does is to facilitate the computation of subjective priors and likelihoods. Lipton suggests that PIBE is a heuristic human beings use to achieve conditionalization is a more efficient way. Both of these variants of compatibilism tries to solve the conflict by focusing on the psychological aspect of Subjective Bayesianism. Weisberg, on the other hand, takes to objective route: he argues that what PIBE does is to characterize some objective *a priori* distribution p. I shall argue, however, that all three attempts are inadequate. I will also use my analysis of these positions as an occasion to explain abductive volunatarism.

2.1 Samir Okasha's Defense of PIBE

Samir Okasha argues for the compatibilist thesis by arguing that there is a more "phenomenologically" accurate representation of how explanation can work along side with conditionalization. His idea is that in actual reasoning, explanatory reasoning can help an agent to carry out Bayesian calculations on the fly. To demonstrate what he means, he asks us to imagine the following scenario: a doctor is trying to diagnose if a child has H_M: pulled a muscle or H_L: torn a ligament. Okasha suggests that her reasoning can be a mix of PIBE and Bayesian conditionalization as follows:

> Suppose we ask the doctor to justify her reasoning. She answers: 'firstly, pre-adolescent children very rarely pull muscles, but often tear ligaments. Secondly, the symptoms, though compatible with either diagnosis, are exactly what we would expect if the child has torn a ligament, though not if he has pulled a muscle. Therefore the second hypothesis is preferable.'[6]

Okasha's suggestion is that the doctor has used to explanatory virtues to arrive at the values for the prior probabilities for the child tearing a ligament or pulling a muscle—call them, respectively, $P(H_L)$ and $P(H_M)$—and the probability of evidence, i.e., the symptoms, given each hypothesis, that is, $P(E_S|H_L)$ and $P(E_S|H_M)$. He argument is that since in this case what PIBE does is not to confer any bonus probabilities on the doctor's degrees of belief after conditionalization, there is no threat of being Dutchbooked. Thus, "the conflict between IBE and Bayesianism alleged by van Fraassen depends entirely on an idiosyncratic way of representing IBE

---

[6]Samir Okasha, "Van Fraassen's Critique of Inference to the Best Explanation," *Studies in History and Philosophy of Science Part A* 31, no. 4 (2000): 703.

in probabilistic terms."[7] His argument, however, is problematic, and does not show what the conclusion he states.

Do we have a genuine case of Bayesian PIBE? It is clear that Okasha intends to use this example to demonstrate how IBE can be used along side of conditionalization, but what is invovled in his example is something quite different. He says that the doctor "has used explanatory considerations as an aid for calculating the priors and likelihoods needed to apply Bayes's theorem itself."[8] Let us recall that the conclusion of a conditionalization is a posterior degree of belief based on the prior distribution and the evidence as expressed by the likelihood distribution. It has to be a numerical value.[9] In other words, conditionalization would get us the actual values of $P(H_L|E_S)$ and $P(H_M|E_S)$. But this is not what the doctor does in both the story and Okasha's analysis. By concluding which hypothesis is more preferable, what the doctor has inferred is a rough estimate of this inequality, $P(H_L|E_S) > P(H_M|E_S)$, or essentially the same expression in the form of a ratio:

$$\frac{P(H_L|E_S)}{P(H_M|E_S)} > 1$$

While this inequality might only look superficially different than the two posterior probabilities, the computations involved are quite different, and has important epistemological implication. The important point is that deciding between two competing hypotheses is a different inferential step than conditionalization. To begin, the actual posteriors would require the following computations:

[7]Okasha, 703.

[8]Okasha, 703.

[9]Or an interval, but we can put this detail aside.

$$P(H_L|E) = \frac{P(H_L)P(E|H_L)}{P(E_S)} \qquad P(H_M|E_S) = \frac{P(H_M)P(E|H_M)}{P(E_S)}$$

Note that E_S, the probability of the child having the symptoms, is required for the calculations of both posterior. To determine this probability, however, we need to know the sum of the probabilities of E_S jointly occurring with each possible hypothesis H_i, assuming that we are dealing with a discrete distribution:

$$\sum_{i=1}^{k} P(H_i)P(E_S|H_i)$$

In other words, we need to know the prior probabilities for the every possible disease the child might have, and the probability of her symptom occurring given this disease. This denominator is sometimes called the *catchall* factor, because it's supposed to catch all the possible occurrences of the data within the sample space. The catchall is computationally demanding, and often intractable. In statistical computing, it is often approximated using simulation techniques such as Markov Chain Monte Carlos (MCMC).[10]

We should then be wary of any example of a human being doing Bayesian calculations on the fly, even if it's an approximation. Wesley Salmon argues that the catchall factor is too computationally demanding for a plausible epistemology. It makes, for instance, any rational reconstruction of scientific reasoning impossible: in order to compute the probability of the occurrence of evidence at hand given the hypothesis in question is not true, it would require us not only to assign probabilities to all of our current theories (an already insurmountably difficult task) but also to have

---

[10]Peter Hoff, *A First Course in Bayesian Statistical Methods* (Springer, 2009) ch. 4.

probability assignments for theories *not yet conceived* (1990, 388). If conditionalization indeed presupposes the need for catchall, its demand is then absurd, for:

> What is the likelihood of any given piece of evidence with respect to the catchall? This question strikes me as utterly intractable; to answer it we would have to predict the future course of the history of science. No one is ever in a position to do that with any reliability (1990, 389).

The same criticism applies to the doctor in Okasha's non-scientific example as well—no amount of explanation is going to help the doctor to figure out what the catchall is. Note that the evidence involved here is that the symptoms expressed by the child. To conditionalize on this evidence, however, require us to know the *unconditional* probability of the child having these particular symptoms. This in turn requires us to know the prior probability of each possible hypothesis H_i and the conditional probability of the symptom E_S on that hypothesis H_i beyond the two hypotheses mentioned by the doctor, unless, of course, all other hypotheses have a zero prior—however I do not think it would be reasonable for the doctor to think that there is an absolute zero probability that the symptoms are due to something other than pulling a muscle or tearing a ligament. Let me be clear that I am not saying that there is no Bayesian or probabilistic analysis of the doctor's reasoning, but I am doubting Okasha's claim that the doctor is using inference to the best explanation to conditionalize on the evidence she's seeing.

There is a reading that allows Okasha to sidestep the problem with catchall, but it will not help his claim that PIBE can be can an ally of Bayesianism. This is because this alternative is to say that the example does not involve conditionalization at all.

The idea is that if we single out two hypotheses for comparison, we can ascertain which hypothesis would have a higher posterior, without needing the catchall. To see this, consider that

$$\begin{aligned}
\frac{P(H_L|E_S)}{P(H_M|E_S)} &= \frac{P(H_L)P(E|H_L)}{P(E_S)} \times \frac{P(E_S)}{P(H_L)P(E_S|H_L)} \\
&= \frac{P(H_L)P(E_S|H_L)}{P(H_M)P(E_S|H_M)}
\end{aligned}$$

In other words, to calculate the ratio of the two hypotheses, we do not need the catchall, since it would just get canceled out. Okasha's example might well be a good intuitive example for this comparative inference, but that's besides the point, since he's supposed to be showing us that PIBE can work with conditionalization, which is required for a Strict Bayesianism. Salmon suggests this as a more realistic approach to Bayesian epistemology, but Earman makes it clear that the downgrade Salmon proposes is not acceptable for any Bayesian, since the lack of actual posterior probabilities has rendered most Bayesian methods unworkable. Thus, while this interpretation of Okasha's example eschews the problem of catchall, it does not show that PIBE can work consistently with conditionalization, which makes the example irrelevant.

It might be argued that Okasha's example could be interpreted even more charitably by supposing that the example is constructed with precise probabilities, so that conditionalization is possible. Still, Okasha's defense appears to me rests on a fundamental misunderstanding of a thesis about prior distributions shared by van

Fraassen and Subjective Bayesians. Neither is committed to idea that priors need to be justified for it to be admissible for conditionalization. Once this is pointed out, the reason for van Fraassen's "idiosyncratic" representation is clear as day—he has no problem with IBE involved with informing the choice of the prior. What he, Bayesians, and the original proponents of IBE are after is how we ought to respond to the deliverance of experience and Bayesians in particular are saying that this is *all* that should matter. This is why van Fraassen's sketch of PIBE is a two-stage process, because as far as he is concerned, the only constraint for the first stage is that your opinion now must be one of the possible opinions you take your future self will hold, i.e., following the principle of Reflection. So as long as your conception of your future self will not take what you consider to be the best explanation now to be impossible, van Fraassen has no problem with it.

Okasha seems to be at least aware of the potential criticism that he might have failed to engage in van Fraassen's argument.[11] His response is that even though his version of PIBE only pertains to how priors and likelihoods are to be arrived, it still figures into the updating process of Bayesianism since the priors and likelihoods are crucial components of conditionalization.

This response, however, is tone-deaf to the iterative nature of Bayesian updating. What makes Bayesian updating so attractive is that not only does it tell us how to revise an ignorant prior opinion, but also how to incorporate evidence into an already informed belief. For each iteration, what was once the posterior distribution would be a prior when new evidence is obtained. Okasha's characterization of Bayesian reasoning completely sidesteps this crucial procedure, since the only case he considers

---

[11]Okasha, "Van Fraassen's Critique of Inference to the Best Explanation," 705–6.

is one where the prior is yet to be informed by the evidence, and this is when the prior could be more or less chosen arbitrarily by the agent without being worried about being Dutchbooked. In other words, during the doctor's first encounter of the situation, she is free to adopt a prior based on her expertise and experience. At this point, she can use IBE to her heart's content within the bounds of logic and probability. For instance, if H_L and H_M are the only two mutually exclusive hypotheses, then the sum of their probabilities cannot be more than 1, but other than that she is under no threat of being Dutchbooked until she revises her belief via conditionalization. If all she does is to compute her posterior probability based on her PIBE-inspired priors and the symptom as evidence, then Okasha is correct in suggesting that she will not be Dutchbooked, since her explanationist musings are contained in the subjective initial prior. But, again, this was never van Fraassen's problem with PIBE.

In other words, a crucial problem never addressed by Okasha is the following: suppose the doctor runs some diagnostic on the child to test her hypotheses, e.g., ordering an X-ray, to gather further evidence, E_x. Now the doctor has to revise her degrees of belief on the two hypotheses based on new evidence. Now, the crucial question is: at this stage, could she revise her no longer ignorant priors beyond what is implied by E_x and Bayes' theorem?

Based on Okasha's idea that PIBE just naturally figures into the agent's phenomenology when making probabilistic judgments, it would appear that the doctor would again reconsider the priors and the likelihoods based on their explanatory virtues. So suppose the answer to the question posed in the last paragraph is *yes*—the doctor should adjust the new priors based on their explanatory virtues beyond what

conditionalization tells us, she would be essentially giving bonus to the posterior probability in a way that makes herself liable to be Dutchbooked as van Fraassen argued. If the answer is *no*, then I don't see how Okasha has engaged in van Fraassen's argument in all, since after all this means that PIBE has no place in the revision of belief in light of evidence. All it does is to influence a prior before undergoing *any* conditionalization, which is already unrestricted for Subjective Bayesians. So in either case Okasha's account is inadequate.

If anything, Okasha's example illustrates a deep tension between the requirement of conditionalization

2.2   Lipton's Argument From Psychological Realism

In broad strokes, Lipton's argument is that explanation serves an approximate realization Bayesian of conditionalization. Lipton's argument shares a similar structure as Okasha's, in that they both are suggesting that in day to day reasoning, IBE-like reasoning is how people think about probability, so Bayesian should want compatibilism as much as explanationists do. Lipton's argument, however, is more empirically informed, since it appeals to the classic literature in behavioral economics done by Kahneman and Tversky. Lipton's aim is also more substantive: not only he claims that Bayesianism and PIBE are compatible, but PIBE in fact complements Bayesianism by providing a psychologically realistic picture of people can approximate Bayesian reasoning in real life.

Lipton's suggestion, then, is more ambitious and explicit than Okasha's: Okasha, for the most part, simply gestures toward the idea that explanation-based inferences

describes how people could come with priors and likelihoods. Lipton's thesis, on the other hand, is much emphatic: he thinks that the behavioral economics and psychology has shown that people use IBE to compute posterior probabilities. Lipton recognizes that this thesis is clashing directly with van Fraassen's argument:

> . . . .it may sound as though explanatory consideration are somehow to modify Bayesian formula, say by by giving some hypothesis a poster 'bonus' beyond what Bayes's theorem itself would grant. . . This is where the spectre of dutch book irrationality appears. . . but it is not what I propose. Instead I want to suggest some of the ways in which explanatory considerations may be our way of running or realizing the mechanism of Bayesian conditionalization—the movement from prior to posterior probability—and our way of handling certain aspects of inference that conditionalizing does not address.[12]

How does that work? The first step of Lipton's argument is that behavioral economics has shown that people are in generally bad at thinking probabilistically, and often draw inferences that violate the axioms of probability. The second step is to claim that these findings support Lipton's theory that people often use IBE to reason about probabilistic problem, even though the result is not always optimal. The last step is to conclude from the previous steps that Bayesianism fares poorly against explanationism as a descriptive theory about human being, so that an irenic approach benefits Bayesianism as much as explanationism. Let us review these steps in turns.

Consider the famous effect noted by Tversky and Kahneman called the *conjunction*

---

[12]Peter Lipton, *Inference to the Best Explanation* (Routledge/Taylor; Francis Group, 2004), 108.

fallacy, which occurs when one judges the probability of a conjunction of two events as being higher than its constituents by themselves. This is fallacious as it violates the basic probabilistic principle: the probability of the conjunction cannot exceed the probabilities of its constituents. Furthermore, Tversky and Kahneman observe that the fallacy can be reliably elicited by a certain form of experimental setting. The paradigmatic case comes to known as the Linda problem. The participants were presented with a description of an individual called Linda:

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

They were then asked to rate the following statements in terms of probability:

1. Linda is a teacher in elementary school.
2. Linda works in a bookstore and takes Yoga classes.
3. Linda is active in the feminist movement. (A)
4. Linda is a psychiatric social worker.
5. Linda is a member of the League of Women Voters.
6. Linda is a bank teller. (B)
7. Linda is an insurance salesperson.
8. Linda is is active in the feminist movement and a bank teller. ($A \wedge B$).

The options that really matter are A, B, and $A \wedge B$. The description of Linda is intended to depict someone who is representative of a feminist, but not a bank teller. As mentioned, however, the probability of the conjunction, $A \wedge B$, cannot exceed either

A or *B*. Tversky and Kahneman reports that an overwhelming number of participants rated the conjunction as being more likely, thus suggesting that human beings are highly susceptible to the conjunction fallacy. Lipton suggests that effects such as the conjunction fallacy establish his premise that people are bad at probabilistic reasoning in general.

The second step of the argument is more contentious. Lipton claims that what the Linda problem shows is that people are using Inference to the Best Explanation as a heuristic to estimate the probabilities. The thought is that people are treating the description of Linda as an event to be explained, and they are simply ranking the statements in terms of their explanatory virtues. According to this interpretation, the subjects are ranking Linda being a feminist banker as a more probable than just a banker, because the former provides a better story that explains the description.

A case for compatibilism emerges from these two premises: the empirical evidence shows that people are explanationists, and not Bayesians. Lipton then argues that, due to this empirical fact, Bayesians should make concession to explanationists so that it could supplement Bayesian's descriptive shortcomings: Bayesian conditionalization is too difficult for people to carry out, so they have to use mental shortcuts such as inference to the best explanation to approximate it.

Lipton's argument for compatibilism rests heavily on his account of human psychology, but why should we accept it? He has certainty given *one* story that would explain the findings in behavioral economics, but this is one of many possible hypotheses. To defuse Lipton's argument, all Bayesians have to do is to point out that Lipton's explanationist account is by no means necessary. Consider, for instance,

the following alternative hypotheses:

1. *representative heuristic*: people often evaluate conditional probability by using similarity as less computationally demanding substitute. So in this story people are rating Linda being a feminist bank teller as being more probably than a bank teller, because a feminist bank teller is perceived to be more similar to Linda's description than a mere bank teller is.

2. *Bayesian confirmation theory*: it has been pointed out that evidence E can confirm more firmly some hypothesis H_1 relative to H_2, even if $P(H_2|E) > P(H_1|E)$. This happens when E raises the probability of H_1 more so than that of H_2, but not enough to exceed it, that is, $P(H_1|E) - P(H_1) > P(H_2|E) - P(H_2)$. For an example, consider a standard deck of cards. Suppose a card is drawn randomly and we are interested in knowing whether it's a picture card (C)or if it is a picture card that is also a spades ($C \wedge S$). We receive evidence E: the card is black. E will raise the conditional probability of ($C \wedge S$), from 3/52 to 3/26, even though it is irrelevant to C, since it goes vacuously from 12/52 to 6/26. According to the confirmation-theoretic analysis, description of Linda E is confirmationally more favorable to $A \wedge B$ her being a feminist banker than B being a banker alone. Thus, people are mistaking the probability of the two statements with how well E confirms them.

So there are accounts for the Linda problem that seem to be just as viable as Lipton's proposal. One of them is even within the Bayesian epistemological framework. Now, I am not trying to argue for one of the alternative hypotheses. What I am suggesting is that Lipton's interpretation is by no means necessary, so his point about

the data fitting explanationism better is unfounded.

Not only do I think Lipton's argument from human psychology unconvincing, but I also believe that his argument misses the mark *even if* we grant his hypothesis, in that Bayesians need help from explanationists, whether or not people in general perform Bayesian reasoning poorly. Clearly people are not naturally Bayesians, in the strict sense of the term—constantly using conditionalization to precisely update their beliefs in an unbounded manner. Such a feat is too computationally expensive and mathematically intractable for machines. But I do not think that Bayesian epistemology has ever pretended to give a descriptive model of human behavior. It is clearly a normative project, so I fail to see why Bayesians ought to be moved by Lipton's argument.

My dismissal of Lipton's argument may sound as though I am motivating an anti-naturalistic and anti-scientific position. Let me be clear that my rejecting Lipton's argument from psychology is *not* to say that epistemology should be contained to the airy heights of ideal normative theories. I disagree with Lipton, not because science is irrelevant, but because I disagree with the idea that a theory of epistemology should aim at a descriptively accurate model of human cognition—armchair epistemology makes poor cognitive science. Still, a plausible epistemology ought to provide *practicable* guidance to any intellectually sincere inquirer's quest of knowledge and rationality. As sincere inquirers, we should recognize that we human beings can make systematic errors in thinking. In this context, the failings of the human mind is relevant insofar as revealing the commons pitfall we should avoid, but this *presupposes* a normative theory, which, I think, should be modeled after the successes of our epistemic endeavors,

instead of our failures. This means epistemology should be modeled after reliable methodologies, such as experimental practices and statistical methods. Thus, my rejection of Lipton's argument is not anti-naturalism; rather, I think he is drawing the wrong lesson from behavioral economics.

3 Abductive Voluntarism

3.1 Voluntarism

What motivates van Fraassen's argument against IBE? It may sound as though he's arguing for a strict Bayesianism by rejecting position that is incompatible with it. This is in fact not true *at all*—van Fraassen is not a Bayesian; his argument against IBE is actually the first step of his argument *against* Bayesian. In this subsection, I will sketch out voluntarism, van Fraassen's alternative to Bayesianism. This will set the stage for the introduction of the position I would like to defend.

Van Fraassen is not a Bayesian; he cannot be clearer about this:

So: I am a probabilist, though not a Bayesian. Like the Bayesian I hold that rational persons with the same evidence can still disagree in their opinion generally; but I do not accept the Bayesian recipes for opinion change as rationally compelling. I do accept the Bayesian extension of the canons of logic to all forms of opinion and opinion change.[13]

My definition of terms as "probabilist" and "Bayesian" deliberately follow van

---

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

Fraassen's usage here, so what he means should be clear to us. He adheres to probabilism in that he accepts a probabilistic analysis of partial beliefs, and that they ought to be constrained by the axioms of probability. His explicit rejection of conditionalization as being 'rationally compelling' amounts to the thought that the revision one's belief via Bayes' theorem is permitted or even encouraged by rationality, it is not a requirement—one does not have to accept it as a mechanical rule to be rational. This makes it clear that his view cannot be regarded as Bayesian by our definition of Strict Bayesianism above. However, he *does* accept the defining feature of Subjective Bayesianism, in that they both allow the reliance of prior opinions without justification.[14] This is why he says that both he and the Bayesians accepts that reasonable disagreement exists, for the same evidence could lead to different conclusions via conditionalization if the priors are different enough.

But if he is not a Bayesian, what is he? Van Fraassen calls his view *voluntarism.* As the quote suggests, voluntarism is a kind of probabilism, but its commitment to the following theses:

Permissive Rationality 'what is rational is precisely what is rationally permitted. Thus we are rational in believing something exactly when we are not rationally compelled to believe the opposite.' [@falsehope 277]

Probabilistic Judgment as a Commitment The expression of my degree of belief is an expression of intention or commitment, and not a mere autobiographical description of one's psychological state. [@beliefwill 251-255]

As we have seen earlier, Subjective Bayesianism already has a permissive attitude

---

[14]Fraassen, 178.

toward our prior beliefs, since one needs not to justify them in order to be rational. However, Subjective Bayesian preserves a sense of objectivity by holding that a rational agent must respond to evidence via conditionalization. So, even if two agents have opposite beliefs about something, the Bayesian hope is that if they diligently update their opinion by using Bayes' theorem, their beliefs will eventually converge to the truth. But van Fraassen's permissive rationality rejects even this. That is, conditionalization is rationally permissible, but not rationally compulsory—one is not required to adopt conditionalization as rule. But van Fraassen's idea of rationality is thoroughly permissive: it's not just that conditionalization is non-compulsory, but that rationality does not require the following of any rule at all.

But why should we think that conditionalization is not rationally compelling? The obvious, albeit disappointing, response is that the burden of proof rests on Bayesians to argue for a strict adherence of conditionalization. There is, however, a power argument by Hilary Putnam that has been referenced by van Fraassen as an object to the Bayesian hope of convergence.[15] Putnam's argument was initially directed toward Carnap's system of inductive logic, but it has been recognized that the same result affects Bayesianism. The gist of Putnam's argument is a *diagonal proof* that no "conditionalization machine" will ever be complete—for any machine T_i that we build, there can always be another machine T_{i+1} that can systematically exploit T_i's learning algorithm in order to lead it to error. In other words, there is always something that a given machine cannot learn, so no Bayesian machine can ever be complete and conditionalization cannot be a complete picture of rationality. Earman

---

[15]Bas C. Van Fraassen, "The False Hopes of Traditional Epistemology," *Philosophy and Phenomenological Research* 60, no. 2 (2000): 260.

argues that Putnam's argument can be blocked by the existence of non-recursive Bayesian learner; however, this also assumes that this non-recursive Bayesian learner can compute the degree of beliefs using some other means, so it seems that the proof still supports van Fraassen's point that rationality cannot just be about following the rule of conditionalization.

Thus, van Fraassen holds that the only normative rationality imposes on us is logical coherence:

> This implies, tautologically, that nothing more than staying within the bounds of reason is needed for this status of rationality—not good reasons, not a rationale, not support of any special sort, not a pedigree of inductive reasoning or confirmation, nothing is needed above and beyond coherence. Thus any truly coherent position is rational.[16]

If coherence is the only thing that holds our rationality together, how, then, can we know that our beliefs are in touch with reality at all? Recall John McDowell's famous characterization of Donald Davidson's coherentism as "a frictionless spinning in the void"[17]. L. J. Savage expresses the same worry in the probabilistic context:

> According to the personalistic view, the role of the mathematical theory of probability is to enable the person using it to detect inconsistencies in his own real or envisaged behavior. It is also understood that, having detected an inconsistency, he will remove it. An inconsistency is typically removable in many different ways, among which the theory gives no guidance for

---

[16]Fraassen, 277.
[17]???

choosing.[18]

In a familiar epistemological jargon, the worry here is that if we let go of both requirements that (A) holding unjustified prior opinions is irrational, and that (B) we should update our belief in light of evidence only using logical rules, then we are left with nothing but a relativism. There will be no normative standard on which we can compare or evaluate two equally coherent position.

Van Fraassen rejects this line of thought vehemently, but if rationality is not about rule-following, then what? The answer is the second thesis of voluntarism stated above: the expression of a probabilistic judgment imposes a normative constraint on one's beliefs. So, even though one is free to maintain any existing belief without justification, the act of accepting that belief has a normative repercussion. This provides a normative standard of criticism for the agent who holds that belief. The idea is that while our inherited beliefs do not owe their rationality to some foundational credentials, it puts a normative constraint on our prospective behaviors. According to van Fraassen's voluntarism, the judgment expressed by reports such "it seems more likely than not that it will rain tomorrow" or "my subjective probability of Duke winning the game tomorrow is about 0.7" are not mere *descriptive* reports of one's internal state; rather, it involves making a commitment

The voluntaristic answer is that we should accept the latter interpretation. A
Deliberative Abductivism

When explanationists cite van Fraassen's anti-IBE argument, it may sound as though van Fraassen takes Bayesianism as the correct position by fiat, so that any

---

[18]Leonard J. Savage, *The Foundations of Statistics* (Dover, 1954), 57.

position that contradicts it is *de facto* an inviable one. Indeed, this is how many explanationists read him. Take Lipton for instance:

> In its simplest form, the threatening argument says that Bayesianism is right, so Inference to the Best Explanation must be wrong.[19]

But given our discussion of voluntarism, this assessment of the situation is not quite right, at least in the original context of the argument. The voluntarist argument is not that it is irrational to use conditionalization or IBE, but that (1) they are not rationally compelling in and of themselves, and (2) repacking IBE as some probabilistic rule is inconsistent with Bayesian conditionalization.

There can be two response to this. The first option is to challenge van Fraassen's negative argument heads on, by defending that both conditionalization and PIBE are rationally compelling. The explanationist arguments we shall examine below take this approach. By taking this approach, explanationists have to argue for the legitimacy of PIBE within the stringent requirement of Bayesianism. To do so, they must give an account of how explanation can influence probabilities without subjecting oneself to a Dutchbook Argument. I shall examine and criticize these arguments in greater details below, but the overall compatibilist strategy is a mixture of neutralizing PIBE so that it will not get in the way of conditionalization, and providing incentives for Bayesians to adopt it by amplifying or emphasizing PIBE in areas where conditionalization comes up short. To anticipate my objections, I will mention that the result of this approach is often an unsavory stew, since it has to dance around neutralizing and strengthening PIBE without either trivializing it or over-promising what IBE can do.

---

[19]Lipton, *Inference to the Best Explanation*, 104.

My alternative proposal is that we should take the path of less resistance: instead of hamstringing explanatory reasoning in order to accommodate the stringent conception rationality of Bayesianism, we should exploit the liberating nature of voluntarism. Like explanationist, I think that explanatory reasoning is an indispensable element in our probabilistic and statistical reasoning, but I also agree with van Fraassen that PIBE cannot be made consistent with conditionalization. To give substance to my position, I will borrow and adapt the conception of abduction found in C. S. Peirce, who was the originator of the term 'abduction'.

Abduction, as I shall use the term, refers to the deliberation involved in the construction of the space of possibilities before induction can take place. More specifically, during the abductive phase of inquiry, in which the space of investigation is delineated by a deliberate process that can involve, for instance, the selection of relevant hypotheses, the choice of a model and relevant parameters for the object of inquiry, sensitivity analysis for prior distributions (if Bayesian methods are chosen), power analysis (for Frequentist methods), etc.

During abduction, explanatory virtue is clearly a relevant consideration, for the construction of a framework pertains to the choice of a hypothesis to be tested, and how much the hypothesis, *if true*, explains would provide ground for making the hypothesis a genuine contender. However, my position disagree with explanationism in that the conclusion of an abductive inference does not have to be the best explanation, nor does it have to be the most probable one. X Thus, the abductive stage is characterized by reasonable disagreements: depending what one values,

For an intuitive example, first consider a non-probabilistic example: the practice

of testing students in an academic example. To be more concrete, consider a logic professor devising a test to probe the understanding of her students.

Suppose Dr.X is the instructor of two logic classes, and is about to give a midterm to both—call them class B and P.

But, in contrast to explanationism, my position does not hold that the conclusion of an abductive inference must be the "best" explanation.

Part of this consists of a choice of hypotheses to be tested

The affinity of my Peircean position and van Fraassen's voluntarism is not accidental, because van Fraassen's position is deeply rooted in the tradition of American Pragmatism, and he repeatedly cites William James' *the Will to Believe* as an important source for his voluntarism. Since my focus in this chapter is explanationism, for the sake of continuity and structure, I shall leave this important connection aside for now, and return to Pragmatism in the next chapter.

What, then, is abductive voluntarism? As implied by the label, it aims to be an extension of voluntarism, with the additional emphasis on *abduction*, which is used here in the way similar to how it was used by Peirce. Contemporary philosophers tend to use IBE and abduction interchangeably. For example, consider Jerry Fodor in his *The Mind Doesn't Work That Way*:

> [My] version [of IBE] claims that the explanation that would, if true, provide the deepest understanding is the explanation that is likeliest to be true. Such an account suggests a really lovely explanation of our inferential practice itself, one that links the search for truth and the search

for understanding in a fundamental way.

Roughly, abductive voluntarism holds that inductive rationality presupposes an abductively chosen framework.

1.

2. Abductive voluntarism maintains vam Fraassen voluntarism's thesis about rationality that it is *permissive*, but would add that the permissiveness is a character of abduction, not induction. This is because once the inductive model has been crafted via abduction, the agent's rationality consists following through the commitments implied in the choice of her model.

It might be objected that objected that the abduction/induction distinction I am borrowing from Peirce has been superseded by or nothing but a rehashing of the one between the contexts of discovery and justification. I do not think that is true, at least not based on how

My proposal is abductive voluntarism: rationality is an iterative process that involves deciding via abduction a local framework in which specific inquiries could be carried out through inductive methods such as Bayesian updating or Frequentist hypothesis testing.

4 Total Evidence and Voluntarism

Keynes, however, is unwilling to wholeheartedly endorse the doctrine of the maximization of utility—he is especially concerned with the fact that expected utility does not

account for the weight of evidence, since utility is only discounted by the magnitude of its probability. So as far as expected utility theorists are concerned, if the two probabilities have the same value, it makes no practical difference if the probability is based on the Principle of Indifference, or a big dataset that is uniformly distributed.

However, Keynes points out that proponents of expected utility theory often *does* implicitly assume the importance of the weight of evidence. Bernoulli, for instance, suggests that rationality demands the utilization of all evidence available to us. This implies, Keynes thinks, that it's always rational to get more evidence, but then it raises another critical question about whether or not one could ever be rational in refusing new evidence.[20] If the answer for the former question is positive, and the latter question negative, then we have to conclude that rationality dictates us that we should never stop looking for more evidence.

Keynes does not make the jump from "using all the evidence" to "get all the evidence" clear. Nevertheless, this problem is revisited many years later in an exchange between Ayer and Carnap. So drawing a parallel between them could be helpful.

In his *Logical Foundation of Probability*, Carnap restates Bernoulli's maxim as "the requirement of total evidence".

> *Requirement of total evidence*: in the application of inductive logic to a given knowledge situation, the total evidence available must be taken as basis for determining the degree of confirmation.[21]

Aver, in response to Carnap, raises the Keynesian question: should "total evidence"

---

[20]John Maynard Keynes, *A Treatise on Probability* (Macmillan; Co., Limited., 1921), 84–85.

[21]Rudolf Carnap, *Logical Foundations of Probability* (Chicago]University of Chicago Press, 1950), 211.

include relevant evidence that I do not yet have in possession?[22] The answer must be "yes", Ayer argues. If finding the truth value of some proposition P could potentially sway the balance of my evidence, then I should definitely acquire it. Thus the principle of total evidence seems to suggest that I am also rationally compelled to consider some evidence I do not yet have.

But Ayer points out that this cannot be the whole picture: taken as a rule of rationality, this means we should never stop acquiring unless we are certain that we have acquired all available evidence. This, however, assumes that we know what evidence is available, but it could often be unrealistic to expect to know how much evidence we *do not* currently have.

Ayer argues that this is only a symptom of a deeper problem about logical nature of probability and its bearing on the rationality of our evidential practice. The logical interpretation of probability, held by both Carnap and Keynes, takes probability as a logical relation between propositions, so within this picture, inductive rationality is a matter of having the right degrees of belief between premises and conclusion—this is analogous to the idea deductive rationality is perceive correctly whether the conclusion follows necessarily from the premises.

Logic, however, does not care about how much evidence we have; it only cares about the relation between our propositions. This problem affects the subjective interpretation of probability as well. As Leonard Savage points out, probability could reveal to us the incoherence within the web of our belief, but it cannot tell us how to resolve it.

---

[22]A. J. Ayer, *Probability and Evidence* (Macmillan, 1972), 56.

According to the personalistic view, the role of the mathematical theory of probability is to enable the person using it to detect inconsistencies in his own real or envisaged behavior. It is also understood that, having detected an inconsistency, he will remove it. An inconsistency is typically removable in many different ways, among which the theory gives no guidance for choosing.[23]

We encountered a version of this problem in chapter 1, in which we considered Keynes' definition of the weight of evidence in terms of the conditional relevance. Once again, the weight of evidence seems directly relevant to inductive reasoning, yet it cannot be easily situated in the probabilistic framework. The notion of resiliency, which we discussed in detail in chapter 1, does not seem to do any better. While it captures the expression of the weight of evidence, it does very little in illuminating on how it dedicates the rationality of our decisions. Both Joyce and Skyrms are silent on this.

Keynes, again, is rather ambivalent about resolving the tension between the weight of evidence and how inductive reasoning is generally conceived. In pass, however, he makes a tantalizing suggestion that the *utility* of evidence could be the component needed to resolve the situation. He begins with the observation that getting evidence for a belief low in evidential weight will "probably produce the greatest amount of good", but the situation is opposite when the evidence for the belief is weighty—"there clearly comes a point when it is no longer worth while to spend trouble"[24] Thus, for a hypothesis of interest H, the same evidence E generates different amount of utility

---

[23]Savage, *The Foundations of Statistics*, 57.

[24]Keynes, *A Treatise on Probability*, 84–85.

relative to the amount of information we already have for H. If an agent has almost no information about H, gathering more information would generate the most utility, but for the same evidence, the demand might to low, because the agent might already have enough information about H, so getting more evidence would yield very little to no utility.

This opens the door to another conception of the weight of evidence, one that is not purely based in the *amount* of evidence, but also on how much how much we *want*. In the *Treatise*, Keynes does switch implicitly these two way of thinking of evidential weight—sometimes refers the weight of evidence a balance "between the absolute amounts of relevant knowledge and of relevant ignorance respectively"[25] In a later chapter of the *Treatise*, he also calls weight "the degree of completeness of the information"[26] These remarks suggest that weight is about how much evidence we desire as much as how much we *do* know. This also allows the possibility that new relevant evidence does not always increase the weight of the evidence, because if I already have enough evidence, getting more might not always be informative.

The puzzle about the utility of evidence, and its bearing on the rationality of the gathering of evidence, has been addressed by Ramsey on an unpublished note. Interestingly, however, he has proven essentially the opposite conclusion reached by Ayer and Keynes: Ramsey shows that we should always look more more evidence, because we can never be worse off from doing it. How can this be?

Ramsey's argument is roughly that, *if* we assume an agent to be a perfect Bayesian and that new information does not cost anything, then she will never be no

---

[25]Keynes, 78.

[26]Keynes, 357.

worse off getting new evidence.[27] In fact, she is guaranteed to be *better* off as long as the new evidence will tell her something new. A perfect Bayesian agent is someone who studiously updates her opinions based on Bayes' rule and then act by choosing the action that maximize her expected utility. Note that this assumes two things: first, for any decision problem she faces, there is always going at least one course of action that maximizes her expected utility, and second, as Skyrms points out, this also implies that the agent knows that she will always *stays* being perfectly Bayesian in the future.

I will make use of an intuitive example rather than reproducing the proof here.[28] Suppose we have three hypotheses about the content of an urn in front of us:

1. H_b: 90 black balls and 10 white balls
2. H_w: 10 white balls and 90 black balls
3. H_n: 50 white balls and 50 black balls.

Suppose we start by assuming $P(H_b) = P(H_w) = P(H_n) = 1/3$—we could have some knowledge that assures us that these are the only three possibilities. There is also a reward of \$1 for picking the correct hypothesis. Our expected payoff for choosing each hypothesis would be the same at $1/3$. Nevertheless, we are allowed to sample with replacement as many times as we wish. Should we get more evidence? Yes, according to Ramsey, we should, and this can demonstrated in terms of an expected utility analysis.

---

[27]F. P. Ramsey, "Weight or the Value of Knowledge," *British Journal for the Philosophy of Science* 41, no. 1 (1990): 1–4, also see @goodtotalevidence and @savage, sec 6.2.

[28]This example is adapted from Isaac Levi, "The Weight of Argument," in *Fundamental Uncertainty: Rationality and Plausible Reasoning*, ed. Silva Marzetti Dall'Aste Brandolini and Roberto Sczzieri (Palgrave MacMilan, 2011), 39–58

To begin, at this point, the probability of getting a black ball is the same as getting a white ball. Let E_b be "a black ball is drawn" and E_w for white balls. So:

$$P(E_b) = P(H_b)P(E_b|H_b) + P(H_w)P(E_b|H_w) + P(H_n)P(E_b|H_n)$$

$$= 1/3(0.9) + 1/3(0.1) + 1/3(0.5) = 0.5$$

And $P(E_w) = 1 - P(E_b) = 0.5$. So, in the event of drawing a black ball from the urn, we would update our belief like so:

$$P(H_b|E_b) = \frac{P(H_b)P(E_b|H_b)}{P(E_b)} = \frac{1/3(0.9)}{0.5} = 0.6$$

Similarly, applying the calculation on the other hypotheses, we get:

$$P(H_w|E_b) = 0.067$$

$$P(H_n|E_b) = 0.333$$

Similar argument can be made by assuming E_w, that is, a white ball is chosen. In that case $P(H_w|E_w) = 0.6$. If we were an ideal Bayesian agent, we should pick H_b if E_b, and pick H_w if E_w. Since an ideal Bayesian would choose the option that maximizes our expected utility, in either case the expected value after drawing from the urn once is 0.6, which is an improvement, since before drawing our expected utility is 1/3 for all options. The net gain in expected utility would be $0.6 - 0.33 = 0.27$, is referred to as *the value of information* in the decision theory literature.[29]

---

[29]Howard Raiffa and Robert Schlaifer, *Applied Statistical Decision Theory* (Harvard, 1964), 89–90.

It turns out that we would be even better off if we were to draw from the urn again. Suppose the first draw yields a black ball. So now we have one piece of evidence in hand. Let us refer to our state of belief after the first draw as $H'_b, E'_b, ..$ and so on. For instance, $P(H'_b) = P(H_b|E_b)$ and $P(E'_b) = P(E'_b|E_b)$. One notable change is that $P(E'_b) = 0.7132$ and $P(E'_w) = 0.2868$. If we draw again and get a black ball, this means:

$$P(H'_b|E'_b) = 0.757$$

$$P(H'_w|E'_b) = 0.009$$

$$P(H'_n|E'_b) = 0.233$$

If a white ball were to be drawn:

$$P(H'_b|E'_w) = 0.21$$

$$P(H'_w|E'_w) = 0.21$$

$$P(H'_n|E'_w) = 0.58$$

Thus, if the second sample is a black ball, we would choose b since it has the maximum expected utility at 0.757, and if we get a white ball, we choose n with the expected value at 0.58. So, the expected utility, if we were to draw from the urn again, is: $0.7132(0.757) + 0.2867(0.58) = 0.706$, which is an improvement over just drawing

---

For a more digestible presentation see @winkler sec.6.3.

once. The net gain is $0.706 - 0.6 = 0.106$. Ramsey's proof shows that we can keep on getting more evidence and we will never be worse off. In fact, we will be better off as long as there is evidence out there we do not yet have.

What should we make of Ramsey's proof? There are two issues involved here. The first is Keynes' observation that evidence can have a diminishing return, so relevant evidence does not increase weight, and the other is how the weight of evidence bears on the rationality of our action, especially when it comes to the gathering evidence. Ramsey's note provides a good answer for the former, but not the latter.

To begin, Ramsey's contribution here is a way for us to think about the relationship between the *weight* of the evidence in possession and the *value* of the potential new evidence.[30] Ramsey clearly thought is the value of evidence E for hypothesis H as something along the line of the difference between the prior expected utility $EU(H)$ and the posterior $EU(H|E)$.

For instance, we saw that in the example above, the posterior expected utility of the first draw was 0.27 higher than our prior expected utility, and we saw a net gain of 0.106 in expected utility if we were to draw again after drawing a black ball, so we see that the first piece of evidence has a higher value than the second one. What Ramsey's proof demonstrates is that new evidence has a diminishing return—I get a "bigger bang for the buck" for my evidence gathering endeavor when I have less evidence. In light of this, Keynes' example of the balance of the evidence unchanged by the introduction of new evidence is then somewhat incomplete. This explains one

---

[30]As noted, this is essentially the idea of the value of information in decision theory, but, as a historical note, Ramsey, inspired by Keynes' puzzle about the weight of evidence, has anticipated this development by many years.

of Keynes' puzzle about worthiness of our endeavor to get more evidence in light of the evidence we have in possession.

However, the broader normative question is still unanswered: how should the weight of evidence guide the rationality of our action? To be sure, I do not question that given some assumptions, Ramsey's result will necessarily follow: the same result is proven by both Good and Leonard Savage, so there is no doubt that the result will holds if the assumptions are granted, but that's a big *if*—we have to question if how often these assumptions actually hold.

The scenario we imagined quickly breaks down once we starts to introduce some sort of cost. It was assumed in the example that it costs us neither money nor time to draw from the urn, but suppose it costs us 25 cents for each sample. This means that we would be gaining only $0.27 - 0.25 = 0.02$ in expected payoff for the first draw, and the second draw would definitely not be worth the additional 25 cents. Or suppose that one dollar is not worthy any endeavor that lasts longer than 15 seconds, and it takes 30 seconds to draw from the urn. As soon as minimally realistic assumptions are introduced, Ramsey's result no longer holds.

Cost might also enter into consideration in different forms, e.g., computational cost or memory. In the same context, Savage ponders over an interesting case that introduces yet another dimension of the problem: consider a very ill person, who is given the option to find out with no cost if the disease she has is mortal. Savage points out that an argument can be made that in this case refusing information could be rational. The thought is that the patient may decide that, based on an assessment of her own personality, she would live the rest of her remaining life in agony if she were

to find out that her disease is very serious, whereas she could live relatively happily without knowing. Savage's response is that in this case the information is not really free; it has a *psychological* cost.[31]

Savage's response was intended to paint a counterexample as an explainable exception to the idea that it's always rational to get free evidence, but this sort of implicit cost is the norm, not an exception: at the minimal, any endeavor to seek more evidence will cost us at least time, and the loss of time is the loss of opportunity. Cost might also enter into consideration in different forms, e.g., computational cost or memory. In light of this, Ramsey's demonstration—the value of evidence depends on how much evidence we have and how much evidence there is to get—does not ease Keynes' concern at all, but it makes the question of evidence gathering more pressing: I might not be getting the most out of my evidential endeavor if it turns out that I could be much better off by examining a different hypothesis, for the very set up for Ramsey's proof requires us to have already chosen *one* hypothesis of interest. But in reality, we often have competing interests. The question is not whether I will get some utility out of gathering evidence for hypotheses in which I have interest, but how I should go about doing it, given the limitation of resource.

Ayer, A. J. *Probability and Evidence.* Macmillan, 1972.

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

Fraassen, Bas C. Van. "The False Hopes of Traditional Epistemology." *Philosophy and Phenomenological Research* 60, no. 2 (2000): 253–80.

---

[31]Savage, *The Foundations of Statistics*, 107.

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

Good, I. J. "On the Principle of Total Evidence." *British Journal for the Philosophy of Science* 17, no. 4 (1966): 319–21.

Hoff, Peter. *A First Course in Bayesian Statistical Methods.* Springer, 2009.

Keynes, John Maynard. *A Treatise on Probability.* Macmillan; Co., Limited., 1921.

Laplace, Pierre-Simon. *Philosophical Essay on Probabilities.* Spring, 1995.

Levi, Isaac. "The Weight of Argument." In *Fundamental Uncertainty: Rationality and Plausible Reasoning*, edited by Silva Marzetti Dall'Aste Brandolini and Roberto Sczzieri, 39–58. Palgrave MacMilan, 2011.

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

Neurath, O., Marie Neurath, and Robert S. Cohen. "Empiricism and Sociology." *British Journal for the Philosophy of Science* 25, no. 4 (1973): 343–52.

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.

Raiffa, Howard, and Robert Schlaifer. *Applied Statistical Decision Theory.* Harvard, 1964.

Ramsey, F. P. "Weight or the Value of Knowledge." *British Journal for the Philosophy of Science* 41, no. 1 (1990): 1–4.

Ramsey, Frank P. *Philosophical Papers.* Cambridge: Harvard University Press,

1990.

Savage, Leonard J. *The Foundations of Statistics.* Dover, 1954.

Winkler, Robert. *An Introduction to Bayesian Inference and Decision.* Probabilistic Publishing, 2010.